

HAVA HARP OKULU
HAVACILIK ve UZAY TEKNOLOJİLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

UÇAK KAZA VE OLAYLARININ VERİ MADENCİLİĞİ YÖNTEMİYLE
ANALİZ EDİLMESİ

Erkan KARACA
Hava Levazım Üsteğmen

ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI

İSTANBUL
Temmuz 2010

Her hakkı saklıdır.

ÖZET

Yüksek Lisans Tezi

UÇAK KAZA VE OLAYLARININ VERİ MADENCİLİĞİ YÖNTEMİYLE ANALİZ EDİLMESİ

Erkan KARACA
Hv. Lv. Ütgm.

Hava Harp Okulu
Havacılık ve Uzay Teknolojileri Enstitüsü
Endüstri Mühendisliği Anabilim Dalı

Danışman : Prof. Dr. Sami ERCAN

Küresel kavramda hava ulaşımının etkin bir şekilde kullanılabilmesi için, uçakların en yüksek güvenlikle idare edilerek, kazalardan korunması gerekmektedir. Bugün havacılıkta en önemli sorunlardan biri de uçuş güvenliğinin artırılmasıdır. Ancak var olan ticari hava yolları sisteminde, her ne kadar teknolojik ve tasarım olarak ileri hava araçları kullanılsa da, insan-makine ilişkisinin olduğu tüm sistemlerin en büyük sorunu olan insan kaynaklı hatalar; birçok uçak olay ve kazaya neden olmaktadır. Yapılan araştırmalara göre de erişilen ortak nokta; havacılık kazalarına % 60 ila %80 arasında değişen oranlarda insan hatalarının neden olduğudur.

Bu çalışmada, farklı birçok alanda başarı ile uygulanan bir veri çözümleme tekniği olan veri madenciliği kullanılarak, 2005 - 2010 yılları arasında Amerika Birleşik Devletleri'nde gerçekleşen sivil uçak kaza ve olayları irdelenmiştir. Çalışmanın amacı; uçak kazalarında insan etmeninin kazalardaki etkinliği arttıran diğer etmenlerle (hava durumu v.b.) olan ilişkilerini inceleyerek uçuş güvenliğini arttırmaktır.

İnsanların kazalara neden olan özelliklerini göstererek anlamlı kurallar elde edilmiş, bulunan kurallarda veri madenciliği sınıflandırma ve birliktelik kuralı yöntemleri ile ayrı ayrı çözümlenmesi sonucu doğruluk ve güvenilirlik bakımından test edilme olanağı sağlanmıştır. Çalışmanın sonucunda sivil havacılık ve devlet kurumlarına kazaların önlenmesi ve uçuş güvenliğinin artırılması hususunda önerilerde bulunulmuştur.

2010, ...103... sayfa

ANAHTAR KELİMELER: Havacılık güvenliği, uçak kazaları, insan hataları, veri madenciliği, sınıflandırma, birliktelik kuralı

ABSTRACT

Master of Science Thesis

ANALYSIS OF AIRCRAFT ACCIDENTS AND INCIDENTS WITH USING DATA MINING PROCESS

Erkan KARACA
Lieutenant

Air Force Academy
Aeronautics and Space Technologies Institutes

Supervisor: Prof. Dr. Sami ERCAN

Airplanes must be managed with high safety and protected from accidents for using air transport effectively in the world. Today one of the most important issues in aviation is increasing flight safety. Although aircraft used in commercial air transport involve advanced technologies and design, human errors, the greatest challenge of human-machine systems, lead to most of the aircraft incidents and accidents. The common point of access according to surveys in literature, as a rough estimate 60% to 80% varying proportions of aviation accidents among the causes of human errors.

In this study, civil aircraft incidents that took place in US between 2005 and 2010 were analyzed using data mining approach. The aim of study is examining relationship between human factors and other contributing factors (weather, etc.) in the accidents to improve the flight safety.

Rules obtained from data about human features that causes accidents with the classification and association rule mining methods. The result of the methods tested from accuracy and reliability standpoints. Consequently potential improvements were proposed to civil aviation and governmental agencies for prevention of aircraft accidents and betterment of flight safety.

2010, ...103... pages

Key Words: Aviation safety, aircraft accidents, human error, data mining, classification, association rule

ÖNSÖZ ve TEŞEKKÜR

Havacılık sektörü her geçen gün büyüyen ve karmaşıklaşan bir yapıya sahiptir. Bu karmaşıklığa yol açan başlıca neden kazalar ve öncüleri olaylardır. Bu yüzden kazalara neden olan etmenlerin farkına varılıp sürekli bir mücadele içerisinde olmak gerekmektedir. Ulu önder M.Kemal Atatürk'ün de dediği gibi "İstikbal göklerde dir." Bu sözden hareketle 20. yüzyılda Türk havacılığı diğer medeniyetlerin her zaman önünde olmak zorundadır.

Bu kapsamda bir hava aracının tasarlanmasından, uçuş aşamasına; bu aracın yaptığı uçuşlardan, bu uçuşların kayıt edilerek çözümlenmesine kadar geçen süreçte yapılacak her türlü çalışma büyük anlam taşımaktadır.

Çalışmalarım boyunca zaman ve mekan kavramını hiçe sayarak değerli yardım ve katkılarıyla bana yol gösteren danışman hocam Sayın Prof. Dr. Sami ERCAN (İstanbul AREL Üniversitesi)'a; yardımlarını gördüğüm Sayın Doç.Dr. Şule GÜNDÜZ ÖĞÜDÜCÜ (İstanbul Teknik Üniversitesi)'ye, samimi desteğini esirgemeyen Sevgili Eşim Canan KARACA'ya teşekkürlerimi sunarım.

Son olarak, veri madenciliği uzun ve zahmetli bir süreçtir. Bu süreçte en önemli bölüm verilerin sağlıklı olarak temin edilebilmesidir. Türkiye'de ulaşım sektörünün ayrılmaz bir parçası olan havacılık alanında verilerin düzenli bir şekilde kayıt edilmesi ve kamuya açık hale getirilmesi bundan sonra yapılacak çalışmalar için büyük bir önem taşımaktadır.

Erkan KARACA
İstanbul, Temmuz 2010

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
ÖNSÖZ VE TEŞEKKÜR	iii
İÇİNDEKİLER	iv
SİMGELER DİZİNİ	vi
ŞEKİLLER DİZİNİ	vii
ÇİZELGELER DİZİNİ	viii
1. GİRİŞ	1
1.1. Çalışmanın Amaçları	5
1.2. Çalışma İçin Kullanılacak Yaklaşım.....	5
1.3. Çalışmada Kullanılacak Veriler	6
1.4. Tezin Organizasyonu	6
2. UÇUŞ GÜVENLİĞİNİN ÖNEMİ ve İNSAN ETMENİ	7
2.1. Giriş	7
2.2. İnsan Etmeni	7
2.2.1. Shel modeli	9
2.2.2. İnsan etmeninin kazalara etkisi	10
2.2.3. İnsan etmeninde meydana gelen kırılma	11
2.3. Hata Modelleri ve Kuramları	11
2.3.1. Tasarıma karşı teknisyen hataları	12
2.3.2. Değişkene karşı sabit hatalar.....	12
2.3.3. Ters çevrilebilir/ters çevrilemez hatalar.....	13
2.3.4. İhmaller, yanılmalar ve hatalar.....	13
2.3.5. Beceri, kural, bilgi tabanlı davranış ve ilişki hataları.....	14
2.3.6. İsviçre peyniri modeli	15
2.3.7. Heinrich modeli	16
2.4. Hatalardan Kaçınma ve Yönetme	17
2.5. Kazaların Oluşumu.....	18
2.6. Uçak Kaza ve Olaylarına Neden Olan Etmenler	19
2.7. Konu ile İlgili Daha Önce Yapılan Çalışmalar	22
3. VERİ MADENCİLİĞİ	26
3.1. Giriş	26
3.1.1. Veri madenciliğinin gelişimi	28
3.2. Bir Veri Madenciliği Projesinde Başarı için Genel Koşullar	30
3.3. Bir Veri Madenciliği Projesini Başarıya Ulaştırmak İçin Dikkat Edilmesi Gereken Hususlar.....	31
3.4. Veri Madenciliğinde Karşılaşılan Önemli Sorunlar.....	32
3.5. Veri Madenciliğinin Kullanım Alanları.....	33
3.5.1. Perakendecilik	34
3.5.2. Bankacılık.....	34
3.5.3. Telekomünikasyon.....	35
3.5.4. Diğer alanlar	35
3.6. Veri Madenciliği Süreci	37
3.6.1. Problemin tanımlanması	37

3.6.2.	Veri seçimi	38
3.6.3.	Veri önişleme	38
3.6.4.	Veri dönüşümü.....	39
3.6.5.	Veri madenciliğinin uygulanması.....	40
3.6.6.	Sonuçların değerdendirilmesi.....	40
3.7.	Veri Madenciliği Teknikleri.....	41
3.7.1.	Sınıflandırma	42
3.7.2.	Kümeleme	48
3.7.3.	Birliktelik kuralı	51
3.8.	Veri Madenciliği Yaklaşımının Çözümleme için Benimsenmesinin Nedenleri.....	52
3.8.1.	Var olan uçak kaza verilerindeki kısıtlamalar.....	52
3.8.2.	Çözümleme için diğerd istatistiksel tekniklerin uygunluğu.....	53
4.	UYGULAMA.....	55
4.1.	Giriş.....	55
4.2.	Analizde Kullanılan Veritabanı.....	55
4.3.	Yazılım Uygulamaları.....	56
4.4.	Uygulamada Kullanılan Algoritmalar.....	57
4.4.1.	Karar ağacı algoritmaları.....	58
4.4.2.	Birliktelik kuralı algoritmaları.....	59
4.5.	Veritabanının Çözümlemesi.....	61
4.6.	Veritabanının Uygulama için Hazırlanması.....	61
4.6.1.	Veritabanının önişlemesi ve hazırlanması	61
4.6.2.	Gürültülü ve aykırı veriler	63
4.6.3.	Kayıp veriler	63
4.6.4.	Verinin dönüşümü	64
4.6.5.	Veri ayrıştırma	65
4.7.	Veri Madenciliği Yöntemlerinin Uygulanması.....	67
4.7.1.	Sınıflandırma	67
4.7.2.	Birliktelik kuralı	72
4.7.3.	Birliktelik kuralı ve karar ağaçları arasındaki farklar.....	75
5.	SONUÇ VE DEĞERLENDİRME	76
5.1.	Genel Sonuçlar	76
5.1.1.	Sonuçların irdelenmesi.....	77
5.1.2.	Yöntemin irdelenmesi.....	77
5.2.	Ticari Havayolu Sektörüne Öneriler	78
5.3.	Kamu Sektörüne Öneriler	78
5.4.	Gelecek için Çalışmalar.....	79
	KAYNAKLAR	80
	EK 1.....	84
	EK 2.....	95
	ÖZGEÇMİŞ	103

SİMGELER DİZİNİ

a_i	Veri Dizisinin Özellikleri
AID	Kaza ve Olay Verileri
ASIAS	Havacılık Güvenliği Analiz ve Paylaşım Sistemi
B (T)	T Veri Dizisinin Bilgi Kazancı
BTS	Amerika Birleşik Devletleri Ulaştırma İstatistik Bürosu
ζ_1	Elde Edilen Kurallar
D	Veri Dizisi
d_k	Sınıflandırıcı Özellik
DTSC	Amerika Birleşik Devletleri Ulaştırma Bakanlığı Güvenlik
E (T)	T Veri Dizisinin Entropisi
EUROCONTROL	Avrupa Hava Seyrüsefer Emniyeti Teşkilatı
$E(\vec{w})$	Yapay Sinir Ağlarındaki Hata Oranı
f	Birleşim Fonksiyonu
FAA	Amerika Birleşik Devletleri Federal Havacılık Kurumu
GYTE	Gebze Yüksek Teknoloji Enstitüsü
HFACS	İnsan Etmenleri Analiz ve Sınıflandırma Sistemi
JSF	Müşterek Taarruz Uçağı
k	'nci Çıkış Nöronu
Libre	0.45359237 Kg.
NAE	Amerika Birleşik Devletleri Ulusal Mühendislik Akademisi
NSC	Amerika Birleşik Devletleri Deniz Kuvvetleri Güvenlik Merkezi
NTSB	Amerika Birleşik Devletleri Ulusal Araştırma Güvenliği Kurumu
OLTP	Çevrim İçi İşlem Süreci
$\ddot{o}_1$	X Dizisinin Özellikleri
s_i	Veri Dizisinin Özellikleri
THY	Türk Hava Yolları
USAF	Amerika Birleşik Devletleri Hava Kuvvetleri
w_{d_k}	Sınıflandırıcı Özelliğin Ağırlığı
X	Veri Dizisi
X^*	X Dizisinin Normalizasyonu

ŞEKİLLER DİZİNİ

Şekil No	Şekil	Sayfa
Şekil 1.1	Kaza sayıları ve uçuş saatlerinin yıllara göre dağılımı.....	2
Şekil 2.1	Shel modeli	9
Şekil 2.2	İnsan etmeninin oynadığı baskın rolün gösterimi	11
Şekil 2.3	Hata zinciri	11
Şekil 2.4	Hedef kalıbı	12
Şekil 2.5	İsviçre peyniri modeli	15
Şekil 2.6	Heinrich piramidi	17
Şekil 2.7	Şirketlerin koruma ve üretim arasında göstermesi gereken denge.	18
Şekil 2.8	Buzdağı modeli	19
Şekil 2.9	5M etmenleri ve arasındaki ilişki	20
Şekil 2.10.	1950-2006 yılları arasında kazaların nedenlerine göre dağılımı ...	21
Şekil 3.1	Veri madenciliği ve bilgi keşfi süreci	37
Şekil 3.2	Veri madenciliği metotları için gerekli zaman	41
Şekil 3.3	Yapay sinir ağlarına bir örnek.....	45
Şekil 3.4	Bir nöronun gösterimi	45
Şekil 3.5	Lojit dönüşümü	47
Şekil 3.6	Lojistik regresyon fonksiyonu	47
Şekil 3.7	Bayes ağı	48
Şekil 4.1	Verilerin ham halinin gösterimi.....	62
Şekil 4.2	C100 verisine ait gürültünün gösterimi	63
Şekil 4.3	Split validate doğrulama aracının veri ile ilişkilendirilmesi	68
Şekil 4.4	Split validate doğrulama aracının gösterimi	68
Şekil 4.5	Ortam 1 veri grubunun karar ağacı görüntüsü	72
Şekil 4.6	Fp-growth algoritmasının veri ile ilişkilendirilmesi	73
Şekil 4.7	Makine veri grubuna ait birliktelik kuralları	74

ÇİZELGELER DİZİNİ

Çizelge No	Çizelge	Sayfa
Çizelge 2.1	Amaç ile ilişkilendirilmiş hata türleri	14
Çizelge 3.1	Veri madenciliği tarihçesi	29
Çizelge 3.2	Veri madenciliğinin kullanıldığı sektörlere göre dağılımı	36
Çizelge 3.3	İstatistiksel çözümleme ve veri madenciliğinin karşılaştırılması ...	54
Çizelge 4.1	Verilerin çalışma esnasında izlediği yol	56
Çizelge 4.2	Veri genel değişimi	65
Çizelge 4.3	Verinin temizlenmesi ve dönüştürülmesinden sonra veri tabanının son hali.....	66
Çizelge 4.4	W-J48 algoritmasının başarı ölçütleri	69
Çizelge 4.5	CHAID algoritmasının başarı ölçütleri	69
Çizelge 4.6	ID3 algoritmasının başarı ölçütleri	69
Çizelge 4.7	Desicion tree algoritmasının başarı ölçütleri	69
Çizelge 4.8	Birliktelik kuralı için kullanılan algoritmaların karşılaştırılması	73
Çizelge 4.9	Birliktelik kurallarının elde edilmesi için seçilen algoritma	74

1. GİRİŞ

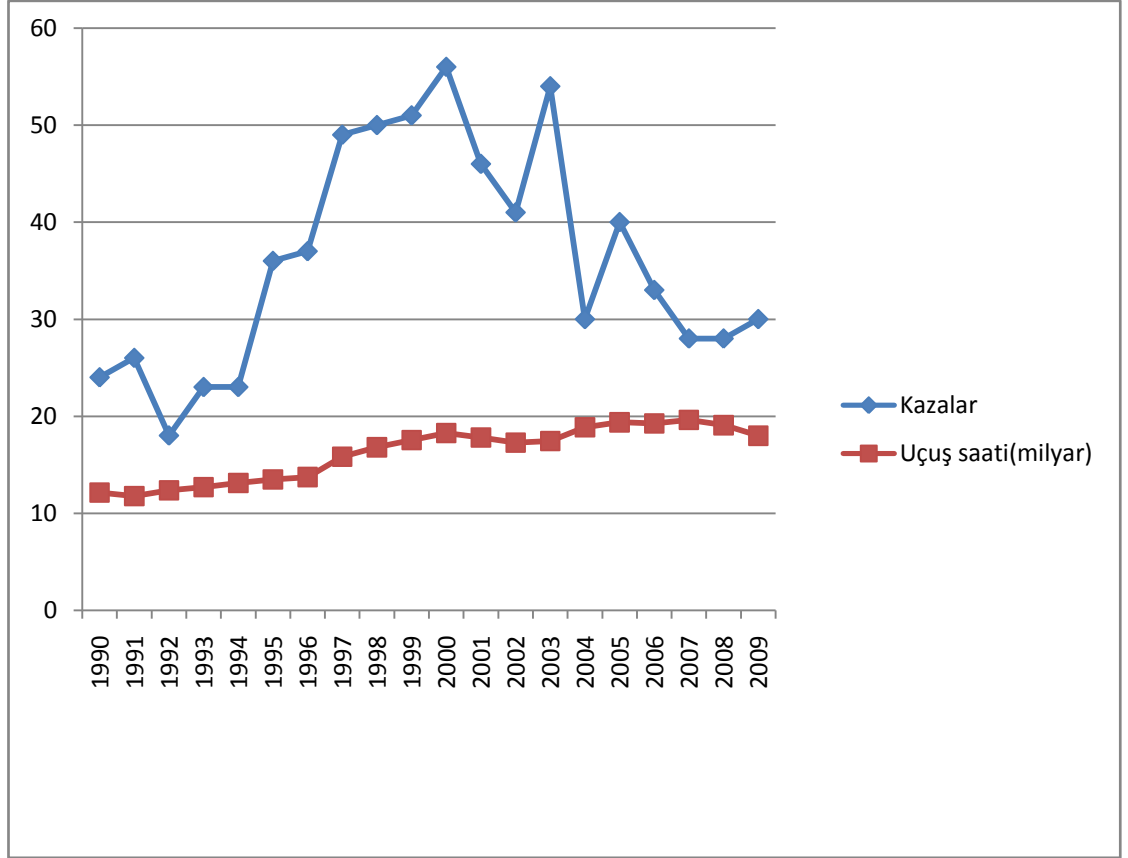
Silas Christofferson'ın, 1913 yılında San Francisco/ Oakland arasını ilk defa deniz uçağı ile yolcu taşıyarak geçişinden beri, bilim adamları yolcu ve kargo uçuşu güvenliğini arttırmaya çalışmışlardır. Peki; zorluklarla ve trajedilerle dolu başlayan bir endüstri, nasıl oldu da bugün en güvenilir ulaşım şekli haline geldi (Shappell vd. 2006)? Bir asır önce ilk yolcu uçuşu gerçekleştiği zaman, kimse bu kadar büyük adımlar atılacağını tahmin edemezdi. Her yerde ve her zaman söylendiği gibi gerçekten de kaza oranı o kadar düşük mü? Veya değil mi? Uçuş güvenliğini etkileyen etmenler nelerdir?

Her yıl dünya çapında 2.271 milyon yolcu ve 41 milyon ton eşya sivil hava taşımacılığı ile taşınmaktadır (Anonim 2008). Sivil hava taşımacılığında var olan potansiyel ve hızlı gelişmelere paralel olarak; uçuş ertelemeleri, uçuştan vazgeçilmesi ve kaza sayıları da kaçınılmaz olmuştur (Gürbüz vd. 2009b).

1990- 2009¹ yılları arasında, uçuş saatleri ve kaza oranları değerlendirildiğinde, (Şekil 1.1.) uçuş saatlerinin sürekli olarak artış gösterdiği, kaza oranlarının da değişken ama çoğu zaman da artan eğim gösteren bir yapıda olduğu görülmektedir. Özellikle kaza eğrisinin düşüşe geçtiği zamanların; Amerikan Ulaştırma İstatistik Bürosu (BTS) ve Amerikan Ulaştırma Bakanlığı Güvenlik Konseyi Başkanlığı'nca (DTSC) yürütülen çalışmalara denk geldiği görülmektedir. Çalışma hava ve diğer ulaştırma hizmetlerinin güvenlik verilerinin kalitesini arttırmak amacı ile veri güvenliği eylem planını kapsamaktadır. Bu çalışmanın hareket planı veri toplama teknikleri, veri bütünlüğü ve analitik teknikleri arttırmayı hedefleyen 10 araştırma projesinden oluşmaktadır. Bilhassa altıncı proje, kazaların öncü nedenlerini belirleme üzerine odaklanmıştır. Planın bir parçası olarak, Amerikan Ulaştırma İstatistikleri Bürosu, kazaların öncü nedenleri ve güvenlik sonuçları arasındaki ilişkileri belirlemeye çalışmıştır. Proje; olası güvenlik problemlerini kestirmeyi ve bunu yaparken de risk-neden ilişkisini göstererek, kaza öncülerini belirlemeye çalışmıştır. Bu çaba sonucunda; kaza öncülerini belirlemenin ve bunun için veri toplamanın ne kadar önemli olduğu görülmüştür. 2003 yılında, Amerikan Ulusal Mühendislik Akademisi (NAE); bir kazadan önce, kaza sinyallerin fark edilebileceğini öngören bir proje üstlenmiştir. Tüm bu çabalara rağmen, düşme eğilimi gösteren kaza eğrisinin, sürekli artan uçuş sayısı ile yukarı yönlü ilerleme potansiyelinde olması doğaldır. Buna karşın kazaları önlemek için sarf edilen her çabanın da olumlu etkileri açıkça görülmektedir (Nazeri 2007).

Amerikan ulusal hava sahası verilerine göre, günlük hava trafik faaliyetlerinde, 174.000 işlem gerçekleşmektedir. Bu kadar büyük rakamlara ulaşan uçuş sayısı için, her zaman güvenli bir havacılık endüstrisi geliştirmek, sektörü zorlamıştır ve zorlamaktadır.

¹ <http://www.nts.gov/aviation/Table2.htm> (15.02.2010)



Şekil 1.1. Kaza sayıları ve uçuş saatlerinin yıllara göre dağılımı

Bu kuruluşlara örnek vermek gerekirse; Amerikan Federal Havacılık Kurumu (FAA) ve diğer sivil havacılık güvenliği kuruluşları gibi. Bu konudaki en büyük sorun nereden başlanacağına karar verilememesidir. Örneğin motorların geliştirilmesi ve ileri gövde teknolojisi, ileri aviyonik¹ ve otomasyon sistemleri başlıkları gibi. Yüzdelerin değişmesine rağmen, yapılan araştırmalar sonucu varılan ortak karar; havacılık kazaları, %60-80 oranında insan odaklıdır. Yani söylenebilir ki: Birkaç istisna dışındaki çalışmalarda; kazalara neden olarak insan hataları yerine daha çok durumsal etmenler veya yalnızca pilotların demografik yapısı incelenmiştir (Nazeri 2007). Hava, görüş, arazi vb. gibi etmenlerin kazalara olan katkısı her ne kadar tartışılmaz olsa da, pilotların veya uçuş sisteminde yer alan insanların, bu koşulları kontrol etkisi çok küçük veya hiç yok gibidir.

Teknoloji ve değişen dünyanın gerekleri ile gereksinimleri karşılamak için hava taşımacılığı büyümek zorundadır. Önümüzdeki 20 yılda, artan yolcu arzının karşılanması için yaklaşık olarak yıllık %5'lik bir büyüme beklenmektedir. Bu da küresel havacılık sistemi için en az iki kat uçak ve uçak hareketi anlamına gelmektedir.

¹Hava araçları üzerinde tüm elektrik ve elektronik sistemleri kapsayan havacılık terimi.

Yani kısacası, gökyüzünde yeni büyük uçakların, talebi karşılamak için kullanılacak olmasıdır. Tüm bunlara ek olarak yolcuların beklentisi de, hava taşımacılığında güvensizliğin azaltılması hatta sıfır ölümcül kaza oranlı bir endüstridir. Hava taşımacılığının göreceli olarak iyi bir güvenlik sicili bulunmasına karşın, kamunun algısı kazalara odaklanmıştır. Fakat yine de hava taşımacılığı yolcularının, 2020 yılına değin yaklaşık olarak üç katına çıkması beklenmektedir. Bu beklenti hava taşımacılığını; kazalara neden olan etmenlerin etkilerini; gelişmiş bilgileri, teknolojileri ve izlekleri daha iyi anlayan yeni güvenlik hedeflerine yönlendirmiştir. İnsan etmeni, uçak tasarımı, inşaat ve bakım araştırılması gereken ana unsurlar olmuştur. Bu ve bunun gibi katkılar sonucunda yapılan kestirimlere göre 2020 yılında ortalama kaza oranı beş kat azaltılması amaçlanmaktadır. Tüm bu çözümlerin beraberinde yeni faydalar sağlaması olasıdır ancak beraberinde yeni riskler de getirmesi kaçınılmazdır (Graaff 2001).

Havacılık güvenliği, küresel kavramda en önemli sorunların başında gelmektedir. Yenilikler ve gelişmelere rağmen bir temel soru hala güncelliğini korumaktadır.” Uçaklar niye kaza yapıyor?” Dürüst olmak gerekirse, cevabın çok da kolay olmadığı görülebilir. Havacılığın ilk yıllarında vuku bulan uçak kazalarında, çoğu zaman uçağın yalnızca hayati olarak pilota ve maddi olarak da yalnızca uçağa zarar verdiği söylenebilirdi. Bu dönemde en ufak bir problemin bile kaza ile sonuçlanmasına “Uçağın hata affetmez.” kuralı neden olmaktaydı. Ek olarak mekanik olarak güvenilir olmasa da belirtilebilirdi. Aradan çok uzun zamanlar geçmesine rağmen modern çağdaki havacılık da aynı ironik olaylara şahit olmaktadır. Fakat modern çağda; uçuş personeli hataları, uçtukları uçaktan daha fazla ölümcül sonuç doğurmaktadır. Yapılan çalışmalar da göstermiştir ki; havacılık kazalarının yine %60-80’i nispetinde, insan hataları ile ilişkilendirilebilmesi mümkündür. Fakat kazalara etki eden tüm etmenleri tamamen uçuş personeline yüklemek; doktora muayeneye gelen bir hastaya, hastalığını ve nedenlerini tespit etmeden “Sen hastasın.” demekle eşdeğer olur.

Peki, niye literatürde kazalar %60-80’inin insan etmeni ile ilişkilendirilmesi yoluna gidilmiştir? Birçok çalışmada insan hataları ile pilot hataları eşanlamlı olarak adlandırılır. Bu yapılırken de uçuşa dahil olan diğer insanlar göz ardı edilmekte ve belki de gerçek nedenler gözden kaçırılmaktadır. Ancak havacılık kazalarını, pilot hataları üzerine ilişkilendirmek yerine bu hataları göstermek ve tümevarım olarak incelemek daha doğru olacaktır. Tüm bunlardan çıkarılabilecek sonuç, kazaların yalnızca bir nedenle veya bir olayla ilişkilendirilemeyeceğidir. Gerçekte kazaların yalnızca bir nedenle incelenmesi çeşitli sorunlarla doludur. Aksine havacılık kazaları bir dizi neden-sonucu barındırır ama en etkin etmen genelde uçuş personelinin tehlikeli hamlesi olur (Shappell ve Wiegmann 2000).

İnsan hatalarının, karmaşık sistemleri nasıl etkilediğinin tespiti, bu tespitlerden hareketle hazırlanmış eğitim programlarının uygulanması, yeni sistemlerin tasarlanması uçak kazalarında insan etmeninin araştırılmasında en önemli amaçlar arasındandır. Kullanılan metotların amacı, hataları karakterize etmek, önlemek ve etkilerini azaltmaktır.

Modern nakliye ve yolcu uçakları, tüm karmaşık tasarım ve üretimine rağmen hala uçuş personelinin yanlış davranışlarının sonuçlarına açıktır. Bu durum tüm uçaklar için geçerli olduğu gibi, insan- makine sistemleri, nükleer santraller, nakliye sistemleri, ileri teknoloji ilaç üretimi ve kimyasal üretim tesisleri için de geçerlidir. Birçok çalışma alanı havacılık ile karşılaştırıldığında daha korkunç kaza oranlarına sahiptir. İnsan hataları çoğunlukla; bakım, yönetim, ortam ve makine gibi ek etmenlerle desteklenir.

Bu çalışmada, hava kaza ve olaylarında, insan etmeninin rolü ve bu etmenin etkisi ile meydana gelen kazaların azaltılması/ önlenmesi için yapılabilecekler araştırılacaktır. Geleneksel olarak hava yolları şirketleri hatalardan korunmak amacı ile eğitim programları, standart izlekleri, kalite kontrol yönetimi, yazılı çek-listler ve kullanma kılavuzlarını tercih etmektedirler. Yine kontrol paneli (bordo paneli) üzerinde olan uyarı ve alarm sistemleri, insan hatasını önlemek için birer nöbetçi olarak bulunmaktadır. Tüm bu alınan önlemlere rağmen, ölümcül hava kazalarının yarısından fazlasına insan hataları neden olmaktadır (Wiener 1993). İnsan, havacılıktaki insan-makine zincirinde en zayıf halkayı oluşturmaktadır. Tasarımdaki ve de aviyonik sistemlerdeki uygunluklar, teknik sistem hatalarını azaltmaktadır. Her ne kadar insan hatalarından kaynaklanan kazaların azaltılması için uçuş personelinin tamamının eğitilmesi konusunda önemli adımlar atılsa da insan hataları halen birçok kazada başrolü oynamaktadır. İnsan hatalarının önlenmesi ve nedenlerinin anlaşılması, havacılık güvenliğinde en başta gelen mücadele konusu olmuştur (Kumar ve Malik 2003).

Son derece açıktır ki; havacılık tarihi, trajik kazalar ve kayıplarla doludur. Ancak 1950'lerin başından beri süre gelen çalışmalar sonucunda kaza oranları azaltılmış ve bugün bir yerden bir yere giderken ticari bir hava yolunu kullanmak, araba kullanarak gitmek veya dünyanın herhangi bir şehrindeki kalabalık yolda yürümekten daha güvenli hale gelmiştir. Yüzyılın başında yapılan ilk uçuştan beri her ne kadar kazalar büyük oranda azaltılmış olsa da, günümüz uçaklarının yolcu ve yük kapasitesinin artması, gerçekleşen kazalarda yaşamsal ve parasal olarak maliyetlerin artması sonucunu da beraberinde getirmiştir. Sonuç olarak askeri ve sivil havacılık kazalarının oranlarının azaltılması için harcanacak çaba her zaman yeni bir anlam taşıyacaktır (Shappell ve Wiegmann 2000).

Bu yüzden ister dünya çapında, ister yerel pazarda faaliyet gösteren bütün uçak firmaları, uçuş esnasında yaşanan olay ve kazaları eksiksiz olarak içeren raporlar hazırlamaktadır. Bu raporlar daha sonra özel ve devlet destekli birçok kuruluş tarafından toplanmaktadır. Bu verilerden incelenerek anlamlı sonuçlar çıkarılabilmesi için önemli yatırımlar ve çabalar harcamaktadır. Bu raporların oluşturduğu veri tabanlarının çok önemli ve değerli bilgiler içermesi şüphesizdir. Ayrıca yine bu raporların bünyesinde, çok karmaşık korelasyonları (bağılılımları) barındırması; incelenebilir olması açısından bir o kadar da uzmanlık gerektirmektedir. Bunu biraz daha açmak gerekirse; verilerin çözümlenmesi de gayret ve zaman gerektirir. Elde edilen veritabanlarının büyüklüğü, her geçen gün veri sayısının artması ve çok sayıda özelliğin bir arada inceleniyor olması bunun nedenleri arasındadır. İstatistiksel veri çözümlemesi, kazaların

nedenlerinin anlaşılmasına olanak sağlar. Buna rağmen birçok olayda, kazalar çeşitli nedenler ve koşullar bir araya gelerek oluşmaktadır.

Bu nedenle bu büyük veritabanlarının incelenmesinde, geleneksel yöntemlerden çok daha yeni ve yetenekli uygulamalara gereksinim duyulması tabiidir. Bu araçlardan biri de karmaşık verilerde gizli kalmış bilgilere ulaşılmasını sağlayan ve sürpriz sonuçlar ortaya koyabilen veri madenciliğidir (Gürbüz vd. 2009a).

1.1. Çalışmanın Amaçları

Bu çalışmanın amacı uçak kaza ve olaylarına neden olan genel etmenler arasında var olan insan hatasının, hem insanın demografik özellikleri- diğer etmenler (makine, ortam v.b.) ile ilişkisini incelemektir. İncelenecek bu ilişkiler kazalarda başrol oynayan insan hatalarının daha iyi anlaşılmasını, insan hatasının başlı başına bir neden olmaktan çok onu tetikleyen nedenlerin neler olduğunu ortaya koymaktır. Böylece, bu nedenlere karşı önlem alınarak, ileride kazalara yol açması da engellenebilecektir.

- 1) Uçak kazalarının karakteristiğini anlamak,
- 2) Uçak kazalarını meydana getiren özellikler ve bu özelliklerin insan etmeni ile olan iç ilişkilerini bulmak,
- 3) Yinelenen uçak kazalarını önlemek, uçuş güvenliğini sağlamak için özel uçak firmalarına ve Türk Hava Kuvvetleri'ne önerilerde bulunmak.
- 4) Veri madenciliği tekniğinde kullanılan metotların ve algoritmaların uçuş verilerine uygunluğunun test edilmesini sağlamak.

Çalışma, bu soruları Amerika Birleşik Devletleri'nde gerçekleşen uçak kaza ve olayları verilerine göre yanıtlamaktadır. Veri madenciliği yöntemi, verileri çözümleyerek aralarındaki iç ilişkileri ortaya çıkarmak için kullanılacaktır.

1.2. Çalışma İçin Kullanılacak Yaklaşım

Çalışma için çok yönlülüğü ile bilinen veri madenciliğinin etki alanı içinde bulunan metotlar, uçak kazalarını çözümlemek için kullanılacaktır. Herhangi bir veritabanını oluşturan etmenlerin özellikleri arasındaki ilişkileri tanımlama kapasitesi ve gelecekte olacak olayların resmini çizebilmesi, veri madenciliğinin en güzel özellikleri arasındadır. Tüm bunları yapmak için veri madenciliği uygulamalarından biri olarak bilinen açık kaynak kodlu “Rapidminer 5.0.6” programı ve yardımcı program olarak “SPSS 15.0” kullanılacaktır.

1.3. Çalışmada Kullanılacak Veriler

Kullanılacak veriler, Amerikan Federal Havacılık Kurumu'nun sivil uçuşlarda 2005-2010 yılları arasında meydana gelen kaza ve olayları içermektedir. Veriler hakkındaki detaylar dördüncü bölümde ayrıntılı şekilde anlatılacaktır.

1.4. Tezin Organizasyonu

- Bölüm 1.1.'den 1.4.'e kadar tezin giriş bölümü sunulacaktır. Ardından tezin yapısı tanıtılacaktır.
- Bölüm 2.'de tezin ana dayanağı olan uçuş güvenliğinin ne kadar önemli olduğu vurgulanacaktır.
- Bölüm 3.'te veri madenciliğinde kullanılan metotlar ve algoritmalar, sonuca ulaşmada verilerin izlediği yol detaylı olarak anlatılacaktır.
- Bölüm 4.'te kullanılan metot ile uçak kazalarının nasıl çözümlendiği ve özelliklerin nasıl test edildiği gösterilecektir.
- Bölüm 5.'te elde edilen kurallara göre çıkarımlarda bulunulacaktır.

Tezin amacına ulaşması için ve çalışma boyunca aşağıdaki terimler tanımlandığı anlamı ile kullanılacaktır:

- Uçak kazası: Hava aracının uçuşu esnasında bir veya daha fazla aracın, uçağın havada veya yerde hasarlanması veya hava aracı içerisindeki veya dışındaki kişilerin ölümüne veya yaralanmasına neden olan olaylardır.
- Uçak olayı: Bir veya birkaç hava aracının uçuşu esnasında, havada veya yerde hava aracının hasarlanmadığı veya hava aracı içerisinde bulunan kişilerin hafif yaralandığı, bir veya birkaç tane ek etmenin de yardımı ile hasara, ciddi yaralanma ve ölüme neden olacak nitelikteki durumlardır.
- Kaza Etmenleri: Uçak kaza veya olayına önderlik eden veya olmasına neden olan olay ve durumlardır.

2. UÇUŞ GÜVENLİĞİNİN ÖNEMİ VE İNSAN ETMENİ

2.1. Giriş

Neredeyse iki bin yıl önce Romalı filozof Cicero, “Hata yapmak insanın doğasında vardır.” sözü ile insanların hata yapma potansiyeline dikkat çekmiştir. Kesin bir gerçektir ki erkek veya kadın bir faaliyete dahil olduğu zaman, insan hatası bir noktada oluşur.

İnsan hatasının en genel tanımı: Planlı bir zihinsel ya da fiziksel etkinlikte amaçlanan sonuca ulaşmada başarısız olma ve bu başarısızlığın şans etmenine bağlı olmadığı tüm durumları kapsayan genel bir kavramdır.

Geçmişte uçak parçaları ve sistemleri güvenilirliği oldukça düşük unsurlardı. Eskiye nazaran şimdiki uçaklar son derece güvenilir olarak tasarlanıp üretilmektedir. Ancak her şeye rağmen günümüzde insan hatası yüzünden bir uçak olayı veya kazası olduğunu duymak normal bir olay haline gelmiştir. Peki, insan hataları kazaların oluşmasında nasıl bir anahtar rol oynuyor?

İnsan hataları nedeni ile meydana gelen uçak kazaları, tüm havacılık kazalarının yarısından fazlasını oluşturmaktadır. İnsan, insan-makine dinamiği içinde hala zayıf halkayı oluşturmaktadır. İnsanın bu dinamik içerisindeki durumunu güçlendirmek için aviyonik tasarım ve güvenilirliğinde ileri çalışmalar yapılarak teknik sistem hataları azaltılmıştır. Ayrıca uçuş personelinin, bu konuda bilinçlenmesi için havacılık kurumları düzenli eğitimlerle personel donanımını arttırmaktadır. Fakat yine de insan hatasının uçak kazalarında önde gelen etmen olmaya devam etmesi engellenememiştir. Bu yüzden günümüzde uçak kazalarında; insan hatalarını anlamak ve önlemek en başta gelen mücadele konusu havacılık güvenliği sorunu haline gelmiştir. Amerikan Hava Kuvvetleri’nde (USAF) meydana gelen 545 uçak kazasının % 52’sine insan hatasının neden olduğu tespit edilmiştir. Bir başka çalışmada Amerikan Deniz Kuvvetleri Güvenlik Merkezi (NSC) çözümlediği 308 uçak olayında, 178’nin (%58) uçuş personelinin hatasından kaynaklandığı tespit edilmiştir (Anonim 2002, Kumar ve Malik 2003).

2.2. İnsan Etmeni

Bu bölümde insan etmeninin ne olduğu ve havacılık endüstrisi için olan önemi açıklanacaktır. Havacılık sektörü insan hatalarını: “Eğer olabilme ihtimali varsa bir gün olur.” ilkesiyle değerlendirmektedir. Havacılık endüstrisinde de diğer alanlarda olduğu gibi insan etmeninin dikkate alınması gerekir. Havacılık sektöründe motorlu bir araç ile uçuşun ilk günlerinde; tasarım, yapı ve uçağın kontrolü uçuşta egemen unsurlardı. İlk

pilotların ana özellikleri, cesaretleri ve yeni üretilen uçakları ikili mücadelelerde ustalıkla kontrol edebilme becerileriydi. Bunun dışında teknik özellikler fazla önemli değildi. Hatta bir uçağın güvenli bir şekilde havalandırılması, çoğu zaman esas hedeflerin başında gelmekteydi. Uçuşta teknik özellikler yavaş yavaş artmaya başlayınca, insanların uçak ile ilgili rolü ön plana gelmeye başladı. Başlangıçta pilotların uçuştaki yükünü azaltmak, uçağı istikrarlı halde tutabilmesine yardım etmek için mekanizmalar eklendi, daha sonra otomatik sistemler uçuş ekibine yol bulma ve iletişim konusunda yardım etmeye başladı. Bu gibi pilotların kabiliyetlerini tamamlayıcı sistemlerin müdahalesi ile havacılıkta insan etmeni olgusu ortaya çıktı. Peki, “İnsan etmeni nedir?” sorusunu açıklamak gerekirse basit olarak; insan hatası, yanılması veya yanılması ile hareketi olarak yorumlanabileceği gibi diğer tanımlamalar da şöyledir: Bazı yazarlar, insan etmenini ergonomi içerisinde de incelemektedir. Bazıları da, insan etmenini bilimsel bir disiplin olarak ve diğerleri de sistem güvenliğine insan katkısının genel bir bölümü olarak görmektedir. Her ne kadar insan etmeninin ‘insanı işe montaj etme ve işi de insana uydurma’ gibi basit tanımları olsa da, iyi bir tanım yapmak gerekirse; insan etmeni, uçuş esnasında insanın yeteneklerinin, sınırlarının değerlendirilmesi ve sistem başarısı üzerindeki etkilerinin incelenmesidir. Bu; uçuş personelinin etkileşimi, sahip oldukları demografik özellikleri, yazılı ve sözlü usuller ve uydukları kuralları ile çevresel koşullardır. İnsan etmeninin amacı, uçuş güvenliğini ve verimi geliştirerek uçuş personeli ile sistem arasındaki ilişkiyi en iyilemektir. Havacılık alanında pek çok farklı şekillerde kullanılan terim, belki de en iyi uçak kokpit tasarımı ve mürettebat kaynak yönetiminde tanımlanmaktadır. Ancak, bu faaliyetler insan etmeni ve havacılık ilişkisinde küçük bir yüzdeyi teşkil ettiğinden açıkça söylemek gerekirse, havacılıkta insan müdahalesi uzun yıllar göz ardı edilmiş ve üzerinde fazla durulmamıştır.

Amerika Birleşik Devletleri’nde (USA) 1988 yılında Aloha Hava Yolları’na ait ve İngiltere’de 1990/Haziran ayında BAC 1-11 yolcu uçaklarında, uçağının ön camının kırılmasıyla gerçekleşen iki kazayla, insan etmeni konusuna keskin bir odaklanma söz konusu olmuştur. Bunun anlamı, insan etmeninin daha önce hiçbir şekilde dikkate alınmaması veya insan hatalarının daha önceki tarihlerde gerçekleşen kazalara etkisinin olmaması değildir. Fakat o tarihten sonra kazalarda insan etmenlerine ve potansiyel çözümlerine dikkat çekilmeye başlanmıştır (Anonim 2002).

İnsan etmeni incelenirken aşağıda sıralanan özelliklerin de dikkate alınması gerekir:

- İnsan fizyolojisi
- Psikoloji (Algılama, biliş, bellek, sosyal etkileşim, hata içerir.)
- Kokpit tasarımı
- Çevre koşulları
- İnsan-makine ara yüzü
- Antropometri (İnsan vücudunun ölçümlerinin bilimsel çalışması)
- Demografik özellikler (Yaş, cinsiyet v.b.)

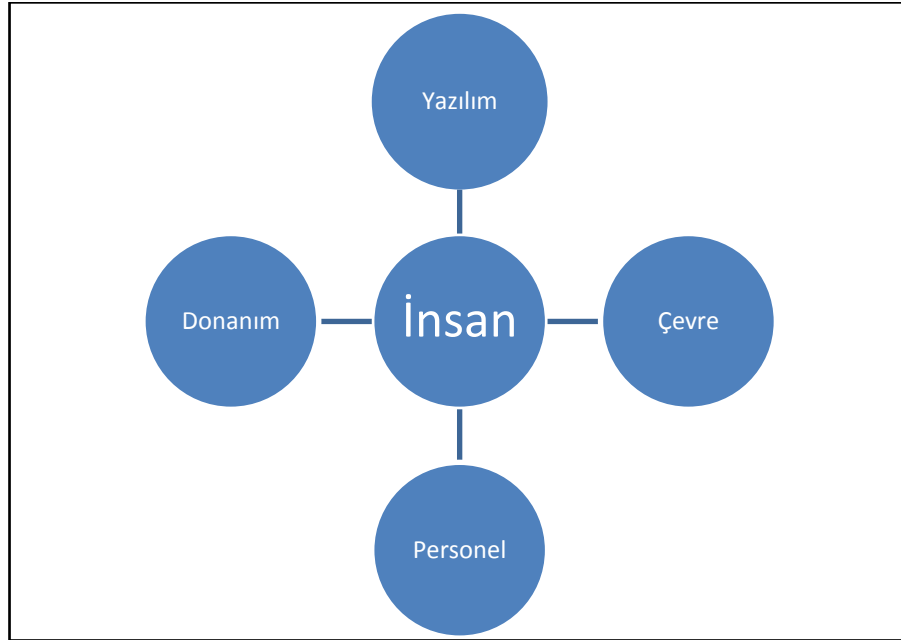
İnsan etmeninin; daha iyi anlaşılabilmesi için kullanılan modellerden biri de “Shel modeli”dir.

2.2.1. Shel modeli (YaDoÇePe)

Shel modeli kısaca, yazılım, donanım, çevre ve personel arasındaki ilişkileri inceler. Bu model, insan etmenlerini daha iyi anlayabilmek için yardımcı olabilir veya insan etmeni konusunda bir çerçeve çizebilir (Şekil 2.1.). İnsan etmeninin bütün diğer etmenlerin ortasında ve birebir hepsi ile ilişkili olduğu görülmektedir.

Modelin baş harflerinden oluşan bileşenleri:

- Yazılım (Uçağın bilgisayar yazılımı, uçuş usulleri, uçuş çek-list'i)
- Donanım (Araçlar, uçağın fiziksel yapısı, uçuş güverte tasarımı, konumlandırma, kontrol ve araçların işletim duygusu)
- Çevre (Uçuştaki fiziksel koşullar, hava durumu, uçuş ortamı, uçuş endüstrisinin toplum tarafından algılanan kısmı)
- Personel (Pilotlar, uçuş görevlileri, uçuş planlayıcıları, hava trafik kontrolörleri, bakım personeli)



Şekil 2.1. Shel modeli (Anonim 2002'den değiştirilerek alınmıştır)

Shel modelini uçuş güvenliği üzerine uygularsak ortaya çıkan aksaklıklar aşağıdaki gibidir.

Yazılım eksiklikleri:

- Usulleri yanlış anlama,
- Kötü yazılmış kılavuzlar,
- Kötü tasarlanmış kontrol listeleri,
- Denenmemiş veya zor kullanımlı bilgisayar yazılımı.

Donanım eksiklikleri:

- Yetersiz araçlar,
- Uygunsuz donanım,
- Kötü uçak tasarım teknikleri.

Çevre eksiklikleri:

- Rahatsız kabin,
- Yetersiz kabin alanı,
- Aşırı gürültü,
- Zayıf aydınlatma.

Personel eksiklikleri:

- Kabiliyet olarak yetersizlik,
- Denetim eksikliği,
- Yöneticilerin destek yetersizliği.

Modern uçakların, günümüzde kendini test etme ve tanı koyabilmesini sağlayan bilgisayar gücüne sahip şekilde tasarlanmasına rağmen, bir şey hiçbir zaman değişmemiştir. Uçaklar; hala insanlar tarafından sevk ve idare edilmektedir. Ancak insanların da sınırları vardır. Shel modeline göre; ortada kalan insanın, diğer etmenlerin (yazılım, donanım v.b. gibi) performansına yardımcı olacak şekilde, sınırlarının dikkate alarak tasarlanması ve adapte edilmesi gerekir. Eğer bu bakış açısını yok sayarsak insan ve uçuş mürettebatı bu durumda yeteneklerini en iyi şekilde sergileyemez, hatalar yapabilir ve de güvenliği tehlikeye atabilir.

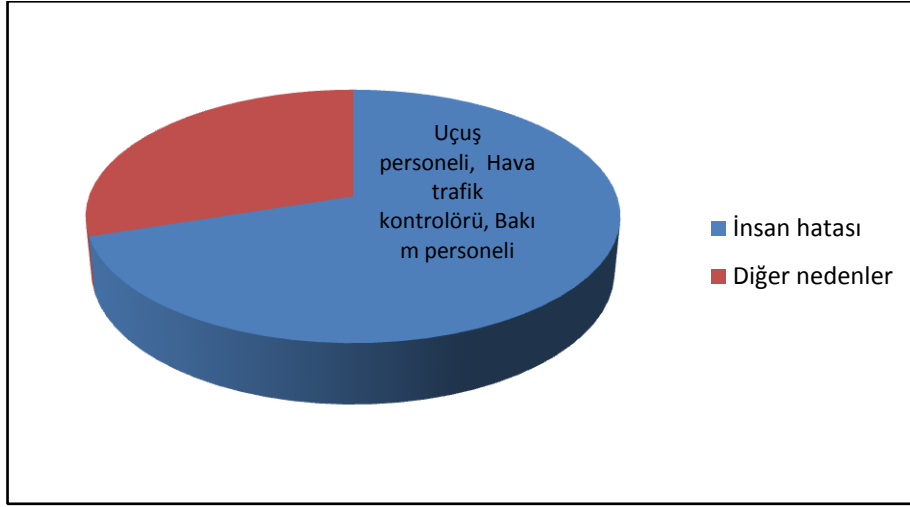
Modern tasarım ve üretim sayesinde uçaklar daha da güvenilir hale gelmiştir. Ancak insanı yeniden tasarlamak olası değildir; o yüzden insanın özünde güvenilmez olduğunu kabul etmek gerekir. Ancak iyi eğitim, usuller, araçlar ve yinelenen denetimlerle bu güvensizliğin bir nebze de olsa aşılabileceği düşünülmektedir. Ayrıca uçak tasarımının en iyi şekilde olması sağlanarak hata potansiyeli de azaltılabilir.

2.2.2. İnsan etmeninin kazalara etkisi

1940 yılında yapılan hesaplama göre, tüm uçak kazalarına insan performansının katkısı yaklaşık olarak % 70 civarındaydı. Uluslararası Hava Taşımacılığı Birliği 35 yıl sonra durumu, insan hataları için yeniden değerlendirdiğinde herhangi bir azalmanın gerçekleşmediğini göstermiştir (Şekil 2.2.).

Elde edilen bütün örneklerde kaza ve olayların önlenemez olduğu ve bazı şeylerin farklı yapıldığı takdirde kaçınılabilir olduğu görülmektedir. Çoğu benzer olayda, eğer bu olaylarda yer alan kişiler farklı hareket etseler ve sorgulasalardı sonuçların çok farklı

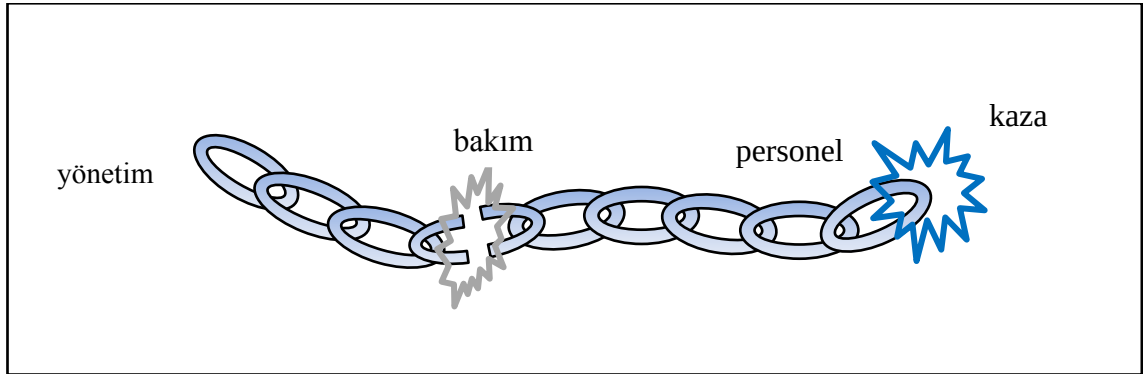
olabileceği gözlenmiştir. Ancak her olayda, bireyler potansiyel tehlikelerin işaretlerini fark edemeyip tepki vermemişler, kendilerinden beklenen hareketi gösterememişlerdir.



Şekil 2.2. İnsan etmeninin oynadığı baskın rolün gösterimi (Anonim 2002'den değiştirilerek alınmıştır)

2.2.3 İnsan etmeninde meydana gelen kırılma

Bütün kaza ve olaylarda bir hatalar zinciri vardır. Bu zincirin son halkası kazadır. Bu halkalardan herhangi biri, bu aşamalardaki sorunlardan bir veya daha fazlası çözülerek kırılırsa bu sayede kaza önlenmiş olur ve güvenlik sağlanır (Şekil 2.3.).



Şekil 2.3. Hata zinciri (Anonim 2002'den değiştirilerek alınmıştır)

2.3. Hata Modelleri ve Kuramları

İnsan hatası; planlı zihinsel ya da fiziksel etkinliklerde amaçlanan sonuca ulaşmada başarısız olma ve bu başarısızlıkların şans etmenine bağlı olmadığı tüm durumları kapsayan genel bir kavramdır.

Hata türlerini değerlendirmek için araştırmacılar birçok model ve kuramlar önermişlerdir. Tüm bu girişimler hatanın doğasını ve karakteristiğini anlamak için yapılmıştır. Bunu göstermek için aşağıdaki model ve kuramlar kısaca açıklanmıştır.

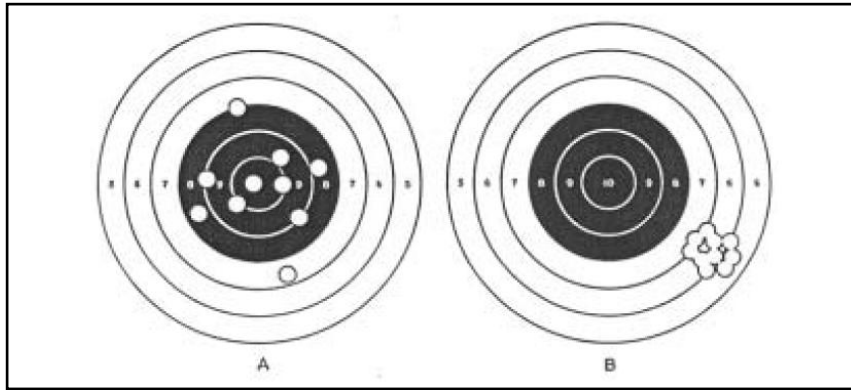
2.3.1. Tasarıma karşı teknisyen hataları

Havacılıkta, hatalar genellikle ön planda bulunan uçuş ekibi, hava trafik kontrolörü ve bakım personeli üzerine vurgulanmıştır. Ancak hatalar, uçak tasarımcıları tarafından uçağın tasarlanması aşamasında yapılmış olabilir. Bu da demek olur ki bir uçak; tasarım kalitesi kadar korunur ve uçabilir. Tasarım esnasında yapılmış bir hata, uçağın güvenli bir şekilde uçuşuna gölge düşürebilir. Alternatif olarak hava yolu şirketi tarafından konulan kusurlu izlekler, bakım organizasyonu veya hava trafik kontrolü yönetimi yine operasyonel problemlere neden olabilir.

Ne zaman bir uçak olayı veya kazası araştırılsa birden fazla hatanın birden fazla kişi tarafından yapıldığı ortaya çıkar. Bunun olması gayet normaldir, çünkü ne zaman kesin bir hatalar birleşimi ortaya çıkarsa o zaman hata savunma duvarı yıkılır ve güvenlik tehlikeye düşer.

2.3.2. Değişkene karşı sabit hatalar

Profesör Reason'a (1990) göre: İki tür insan hatasından söz edilebilir; değişken ve sabit. Şekil 2.4.'te A' da değişken hatalarının doğada rastgele olduğu, B' de sabit hataların tutarlı sistematik bir yol izlediği görülebilir. Burada ima edilen sabit hataların tahmin edilebilir ve bu nedenle de kontrol edilebilir olduğu, değişken hataların kestirilemez ve anlaşılması çok daha güç olduğu görülebilir. Eğer görevin ve bireyin doğası hakkında, gerçekleşen çevre, performansı idare eden mekanizmalar ile ilgili bilgi varsa hatayı kestirmek için büyük bir şans da vardır (Anonim 2002).



Şekil 2.4. Hedef kalıbı (Anonim 2002)

Şekil 2.4.'te de görüleceği üzere; hedef kalıplara, iki tüfekçi tarafından on defa ateş edilmiştir. Tüfekçi A'nın hedef kalıbı sabit hata göstermemekte aksine büyük değişken hatalar içermekte; tüfekçi B'nin hedef kalıbında büyük sabit hatalar içermekte ama küçük değişken hatalar içerdiği görülmektedir. Daha basit açıklamak gerekirse; tüfekçi A'nın her atışında birbirinden farklı isabetler yapması, her defasında farklı hata etmenlerinin olduğunu göstermektedir. Ancak buna karşın tüfekçi B'nin hedefi bulamaması ve tüm atışların yalnızca bir noktada buluşması her defasında aynı hatanın yapıldığını göstermektedir. Yani; tüfekçi B'yi kestirmek ve düzeltmek potansiyel olarak daha kolaydır.

Ancak, kesin kestirimlerde bulunmak için yeterli bilgiye nadir olarak sahip olunur; örnek vermek gerekirse, aynı uçakta iki gün içersinde beş defa uçuş yapan bir pilotun, personel istirahati kullanmadan başka bir uçuşa çıkması, onun kaza yapabileceği tahminini doğurabilir. Bu tarz kestirimleri daha fazla bilgi ile yapmak olasıdır ama her zaman rastgele hata ve elemanlar ortaya çıkabilir ve yanıltıcı olabilir.

2.3.3. Ters çevrilebilir/ ters çevrilemez hatalar

Hataları kategorize etmek için geri dönülür veya geri dönülmez olduğuna karar verilmelidir. Birincisi geri kazanabilir fakat ikinci genellikle olmaz. Örneğin bir pilot taşıyabileceği yakıtı yanlış hesaplırsa, en yakın hava alanına inebilir fakat yanlışlıkla yakıtı havada depolardan boşaltırsa, bunu çözecek hiçbir seçeneği olmaz.

2.3.4. İhmaller, yanılmalar ve hatalar

Burada hatanın doğasına dikkat çekerken amaç kavramı, sorular sorarak vurgulanmaktadır;

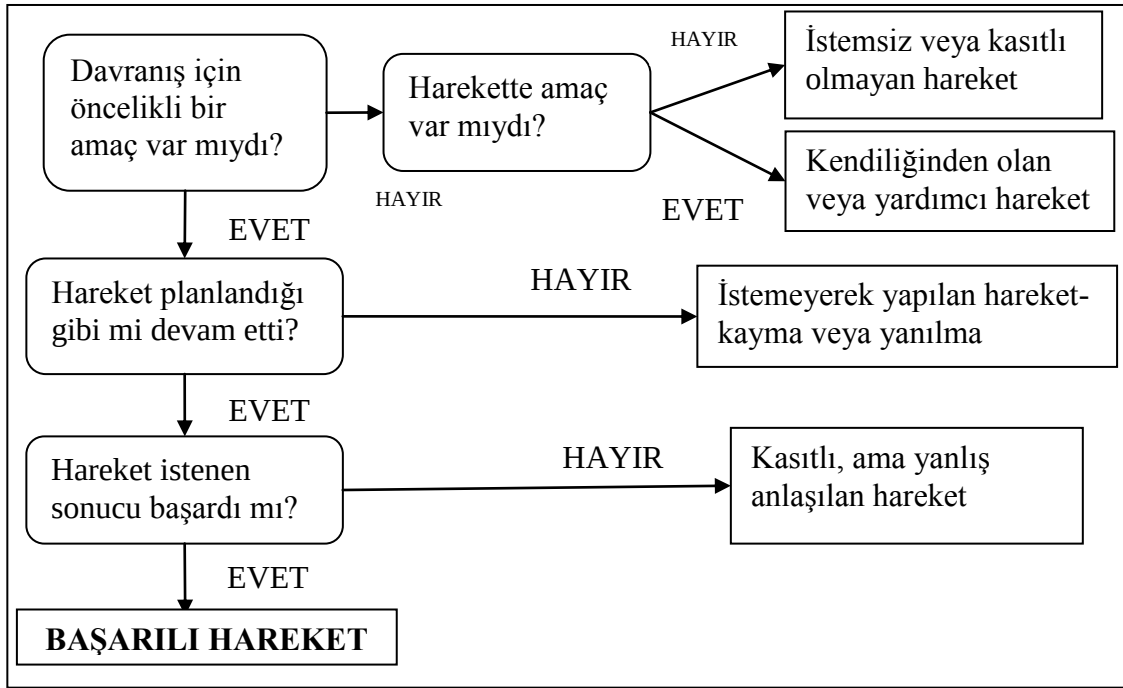
- Eylemler bazı öncelikli amaçlar tarafından mı yönetiliyor?
- Eylemler planlandığı gibi mi devam etti?
- Eylemler beklenen başarı ile mi sonuçlandı?

Bunu daha iyi anlayabilmek için çizelge 2.1.'deki izlek takip edilebilir. Daha sonra çizelge 2.1.'de gösterilen soruları yanıtlamak için hata sınıflandırması önerilebilir.

İhmaller (gözden kaçınma); amaçlanmış veya planlanmış eylemler değildir. Örnek olarak; rakamları kopyalarken yerlerini değiştirmek veya bir izleğin basamaklarını atlamak verilebilir.

Yanılma; eksik yapılan eylemlerdir. Bir kişinin bir eylemi yaparken unutkanlık ya da dikkatsizlik nedeniyle hata yapması örnek olarak gösterilebilir.

Çizelge 2.1. Amaç ile ilişkilendirilmiş hata türleri (Anonim 2002’den değiştirilerek alınmıştır)



Hatalar; yanlış bir amaçla/ planla yapılan özel bir hata türüdür. Doğru olduğunu düşündüğü fakat gerçekte yanlış olan bir eylemi yapmadır.

İhmaller tipik olarak görevin uygulama aşamasında, yanılmalar bilgilerin saklanması aşamasında, hatalar planlama aşamasında gerçekleşir.

İhlaller (kuralsızlıklar) bazen insan hataları olarak görünür; ancak ihmaller, yanlışlar ve yanlışlardan farklıdır çünkü kasıtlı olarak yapılan doğruluğa göre kabul edilmeyen eylemlerdir. Örnek olarak bir pilotun hava alanına daha erken varmak için olması gerek irtifadan daha alçak irtifaya inmesi gibi.

2.3.5. Beceri, kural, bilgi tabanlı davranış ve ilişki hataları

Uçuş personelinin davranışları üç farklı kategoriye bölünmüş olabilir: Kural ve bilgiye dayalı, davranış tabanlı, beceri tabanlı.

Kural tabanlı davranışlar kendisi için rutin veya daha önce öğrenilmiş olan izleklerdir. Bu kuralın bileşenleri ayrı beceri dizileri oluşturabilir.

Bilgiye dayalı davranışlar izlekleri içermez. Bu kural bilgileri değerlendirmek ve kendi bilgisi ve deneyimi ile durumla başa çıkacak bir plan yapmaktır. Bu tür davranışların

her biri özel hatalarla ilişkilidir. Beceri tabanlı hatalara örnek olarak ihmaller, tersine dönme ve çevrenin etkisi gösterilebilir.

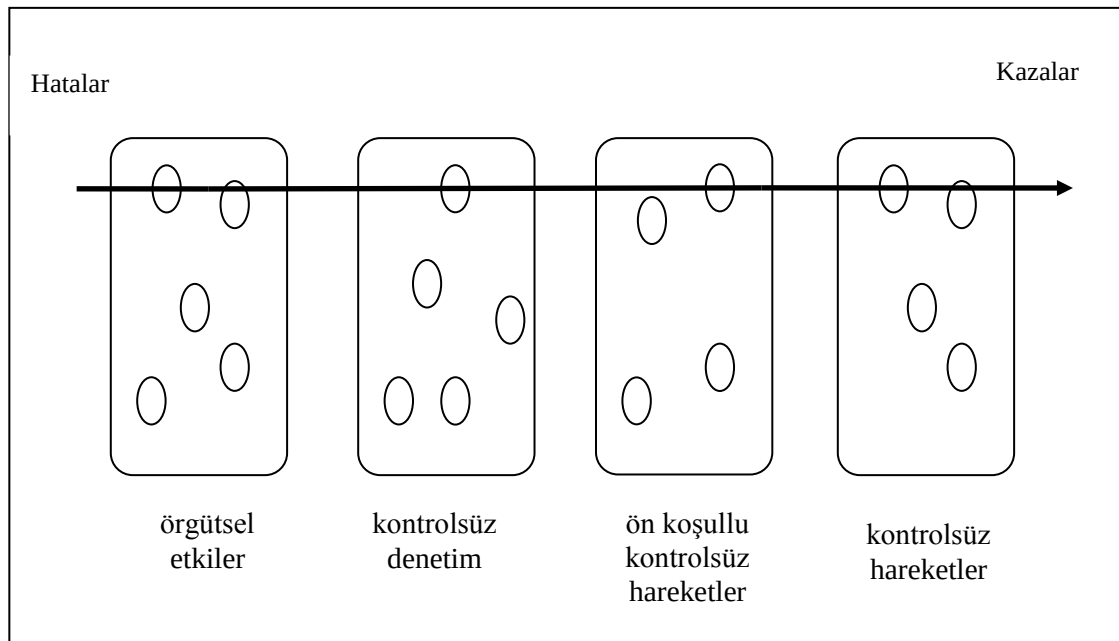
2.3.6 İsviçre peyniri modeli

Havacılık güvenliği konusunda, en etkili kuramlardan biri Profesör Dr. James Reason (1990) tarafından tanıtılan İsviçre peyniri modelidir. Bu model insan hatasını dört düzeyde de tanımlar ve her birinin bir diğerini etkilediğini vurgular.

Bu seviyeler:

- Örgütsel etkiler,
- Kontrolsüz denetim,
- Ön koşullu kontrolsüz hareketler,
- Kontrolsüz hareketler.

Tehlikeli bir durum, kaza haline gelmeden İsviçre peyniri dilimlerinden başarılı bir şekilde görülebilir. Bazı hatalar gizlidir, bunun anlamı geçmişte aynı noktada yapılmış ve uyku durumunda beklemektedir. Bu bir uçağın tasarımı esnasında ya da bir yönetim kararı ile ilişkili olabilir. Her bir dilim hasarı önlemek için tasarlanmıştır. Ancak, her dilim de sistemin veya sürecin hatalarını sergileyen delikler vardır. Her hata sıklıkla zararlı sonuçlar doğurmadan oluşabilir ancak ne zaman kombine hale gelir yani delikler bir hat halini alır o zaman kaza oluşur (Şekil 2.5.).



Şekil 2.5. İsviçre peyniri modeli (Nazeri 2007’den değiştirilerek alınmıştır)

Profesör Dr. Reason (1990) araştırmasında; bir organizasyon ile insan hatalarına karşı savunma kavramını vurgulamıştır. Örnek vermek gerekirse; yinelenen denetlemeler,

pilotların uçuş öncesi fonksiyonel testleri v.b. insan hatalarının tuzaklarını engeller, olası negatif sonuçlarını azaltır. Ne zaman bu savunmalar zayıflar ve kırılır o zaman insan hataları kazalar ve olaylara neden olur. Bu savunmalar diyagramlar şeklinde resmedilmiştir, tıpkı birkaç dilim şeklindeki İsviçre peyniri gibi (Şekil 2.5.). Bundan dolayı bu modelin ismi Profesör Dr. Reason'ın İsviçre peyniri modeli olmuştur (Nazeri 2007).

İsviçre peyniri modeli, yaygın bir kalıp olarak insan etmeninin araştırılmasında kullanılmaktadır. Bu yüzyılda Shappell ve Wiegmann, Profesör Reason'ın kuramlarını alarak "İnsan Etmenleri Analiz ve Sınıflandırma Sistemi'ni (HFACS) geliştirmişlerdir. İnsan Etmenleri Analiz ve Sınıflandırma Sistemi, Profesör Dr. Reason'un insan hataları ile ilgili söz ettiği dört düzeyden başka; onbeş adet başka nedensel kategoriye içermektedir. Bu geliştirdikleri çerçeve; askeri ve genel havacılığa uygulanmıştır

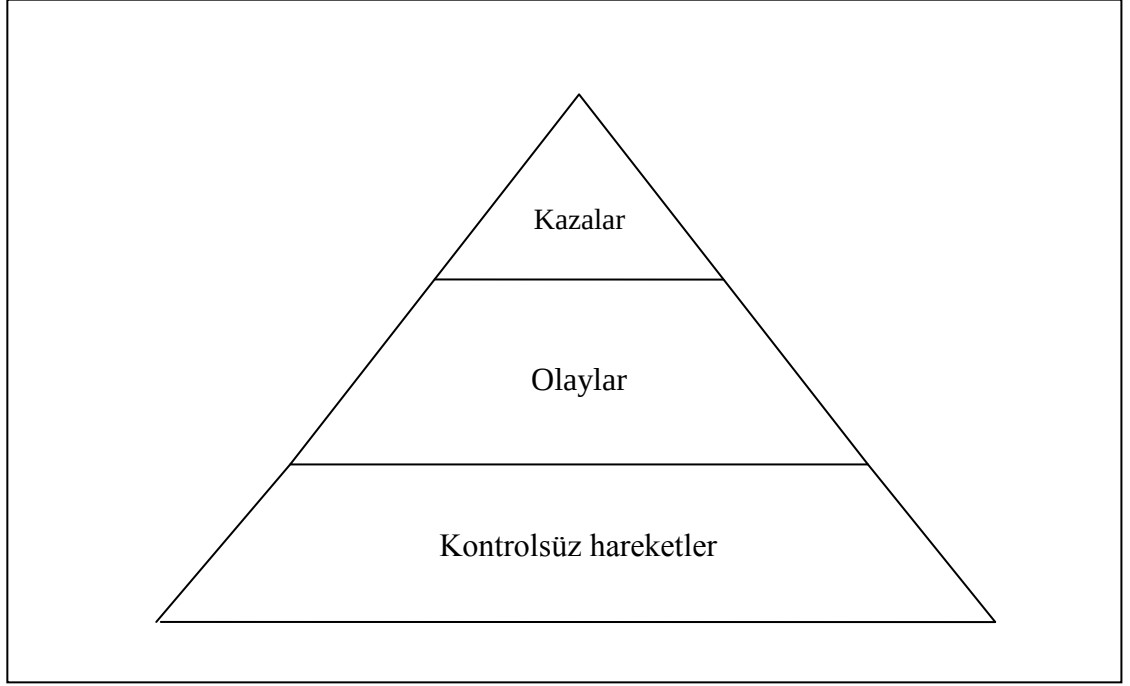
2.3.7 Heinrich piramidi modeli

Geçmiş yıllarda bir başka etkili kuram endüstriyel kazaları önlemek için H.W. Heinrich (1931) tarafından ortaya atılmıştır. Heinrich kuramına veya Heinrich piramidine göre; kazaların nedenlerinden düşük frekanslı ve yüksek riskli olaylar (Şekil 2.6.'da görüldüğü gibi) piramidin en üzerinde yer alır. Piramitte aşağı doğru gidildikçe, sonraki tabaka daha az tehlikeli fakat yüksek oranlı oluşan uçak olayları ve güvensiz hareketlerinden oluşur. Bu kuram uçak olaylarını ve daha az ciddi durumları aralarında istatistiksel bir ilişki olduğunu kabul ederek kazalara bağlar. Piramit her 300 güvensiz harekette, 29 küçük yaralanma ve büyük yaralanma olduğunu göstermektedir. Heinrich piramidini havacılık güvenliğine adapte ederken şu gerçek ortaya çıkmıştır; her büyük kaza öncesinde 3-5 ölümcül olmayan kaza, 10-15 uçak olayı ve yüzlerce raporlanmayan olayı içerir.

Hataların oluşması ile sorumlu olan yönetim fonksiyonları üzerine odaklanan İsviçre peyniri modelinin aksine, Heinrich kuramı hataları, kullanıcı yanlışlarına dayandırır ve bu nedenle güvensiz davranışlar kontrol altına alınarak kazaların azaltılabileceğini önerir (Nazeri 2007).

Kazaların, insanların güvenilmez davranışlarının kontrol altına alınarak azaltılması yaklaşımı, bu çalışma için öneri olarak alınmıştır. İnsanların güvenilmez davranışları da yine insanın kendi özellikleri üzerine çözümlemeler yapılarak engellenebilir. İnsanların demografik özellikleri çözümlenerek geri beslemelerle kazalar azaltılabilir.

Örneğin belirli yaş aralığında yüksek kaza frekansı görülüyorsa, o yaş grubu yeniden eğitim altına alınarak kaza yapma riski azaltılabilir. Yahut o yaş grubu, mesleğin son yıllarında emekli edilerek/ iş akdine son verilerek riskli yaşa mensup şahıslar sistemden çıkartılabilir.



Şekil 2.6. Heinrich piramidi (Nazeri 2007'den değiştirilerek alınmıştır)

2.4. Hatalardan Kaçınma ve Yönetme

Havacılık sektörü, her zaman hatalar oluşmadan önce engellemeyi amaçlamaktadır. Çünkü tamamen ortadan kaldırmak hiçbir zaman olası olmayacaktır. Bu nedenle tüm kuruluşlar hataları yönetmeyi amaçlamıştır.

Hata yönetiminin amaçları;

- a. Hataları oluşmadan önlemek,
- b. Hataların kötü etkilerini ortadan kaldırmak/ azaltmaktır.

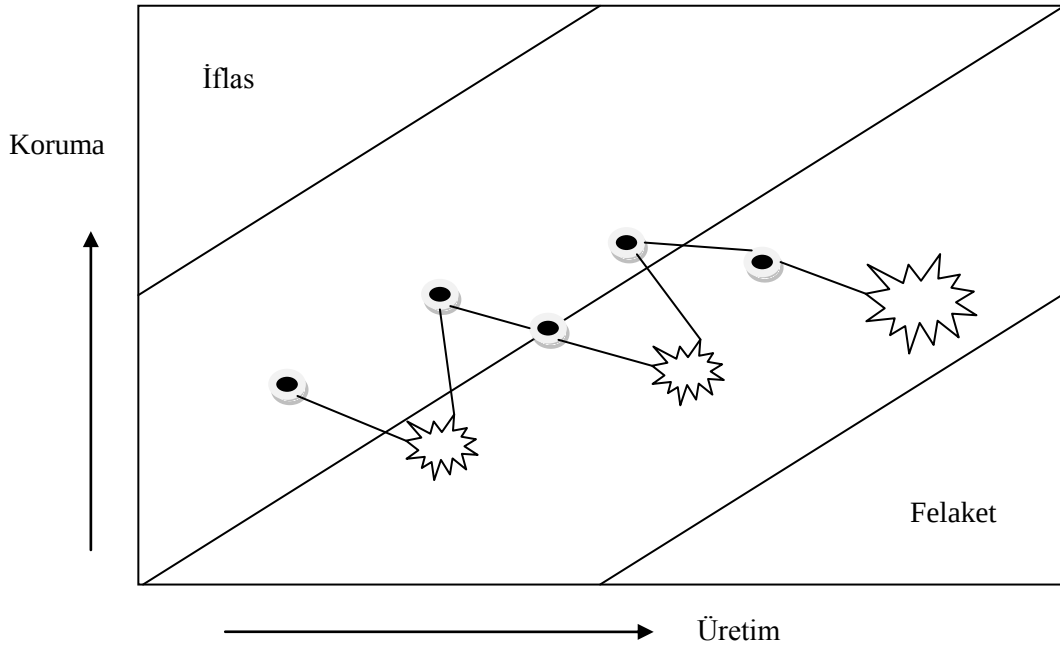
Hataları oluşmadan önce önlemek için, onların en çok nerede olduğunu kestirmek ve sonra yerine önleyici tedbirler koymak gerekir. Veri madenciliği kestirim yöntemleri işte bu görevi üstlenebilir. Bir uçuş organizasyonunda oluşturulacak kaza ve olay verilerinin saklanacağı güvenlik yönetim sistemi; potansiyel zayıf noktaları, hata eğilimli faaliyetleri ve durumları sağlayan mekanizmaların tanımlanmasını sağlar. Bunu da ilk elden yaptıkları hatalarla geri dönülemezliği sağlayan uçuş personeli üzerine uygulayabilirler. Bunun sonucunda da yerel eğitimler, firma izlekleri, yeni savunma düzenleri veya var olan savunma sisteminin değiştirilmesi gibi eylemler yapılabilir.

Hata yönetimi şu gibi tedbirleri içermelidir:

- Birey veya takımın hata sorumluluğunu en aza indir,
- Özel belirli görevlerin veya görev elemanlarının hata açığını azalt,
- Keşfet, değerlendir ve sonra hata/ ihlal üreten etmenleri işyerinden çıkar,

- Birey, takım, vazife veya işyerinin içinde hata/ihlal üreten öğeleri yaratan organizasyonla ilgili etmenleri teşhis et,
- Hata algılamasını arttır,
- İşyeri veya sistemin hata toleransını arttır,
- Sistemi çalıştıran ve idare eden gizli koşulları daha görünür yap,
- İnsanın yanılabilir olmasından dolayı organizasyonun iç direncini geliştir.

Sonuçta, havayolu organizasyonları hataları önlemek, azaltmak ve tespit etmek için tedbirler almak ve kazanç elde etmek arasında uzlaşması gerekir. Bazı tedbirler az maliyetli, bazıları yüksek maliyetlidir. Uçak olayları kısa vadede hata azaltıcı tedbirlere yönlendiren sonuçlar içerir, ama bir organizasyonda uzun zamandır uçak olayı olmuyorsa ya da oluyor da hakkında bilgi sahibi olunmuyorsa, önemine değer verilmiyorsa orada rahatlıktan dolayı bir tehlike vardır ve hatalara karşı savunma stratejilerinde maliyet azalması vardır. Bu şekil 2.7. ile görülebilir;



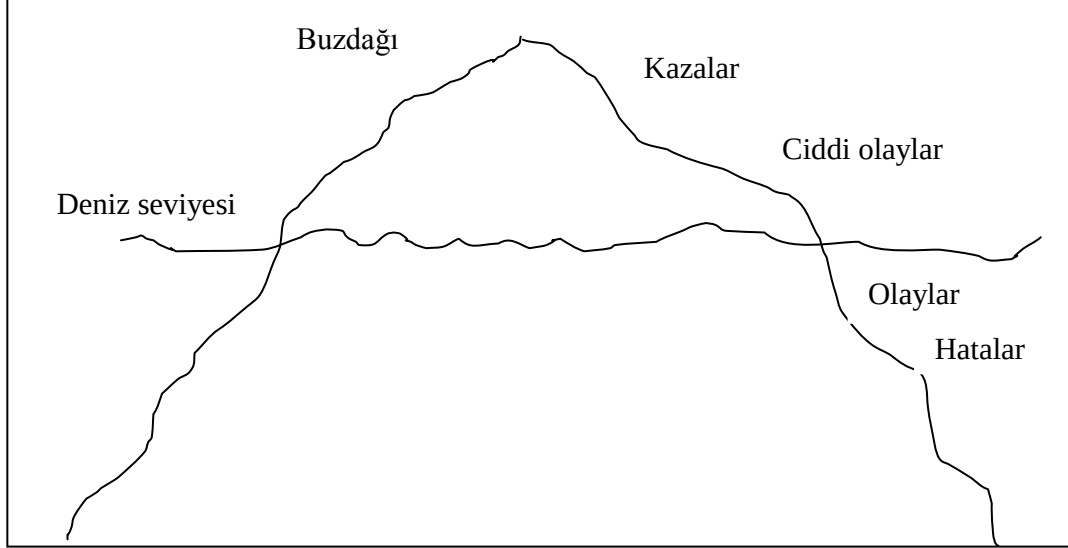
Şekil 2.7. Şirketlerin koruma ve üretim arasında göstermesi gereken denge (Anonim 2002'den değiştirilerek alınmıştır)

Organizasyonların kazanç-maliyet dengeleri, hangi savunucularının hataları yakalamak ve felaket getirecek sonuçları engellemek için etkin maliyetle nereye konması gerektiği önemlidir.

2.5. Kazaların Oluşumu

Kötü durumlarda, uçuş ekibinin yapacağı insan hataları uçak kazalarına neden olabilir. Ancak şekil 2.8.'de tasvir edildiği gibi, kazalar hataların gözle görülen belli olan kısımlarıdır. Bu bir buzdağına benzetilirse, kütesinin en büyük kısmının suyun altında

kaldığı görülebilir. Hataların çoğunluğunun da gerçekte kazalarla sonuçlanmadığı görülür. Ancak birçok hatanın bir araya gelerek olayı; yine birçok olayın da bir araya gelerek kazaları meydana getirdiğini düşünülecek olursa “Buzdağı Modeli” geçerliliğini göstermiş olur.



Şekil 2.8. Buzdağı modeli (Anonim 2002’den değiştirilerek alınmıştır)

Kazalara neden olmayan hatalar uçak olayı olarak bilinen problemlere neden olur. Bunların bazıları (uçuşta önemli hadiselerle neden olan hatalar, tesadüfen ve ya pilotun becerisinden dolayı kazanın oluşmadığı gibi) diğerlerine göre yüksek görüntü gösterirler. Ancak bütün uçak olayları uçak bakım endüstrisi için önemlidir çünkü bütün uçak olayları gelecekte farklı durumlarda olacak potansiyel kazaların uyarıcısıdır.

2.6. Uçak Kaza ve Olaylarına Neden Olan Etmenler

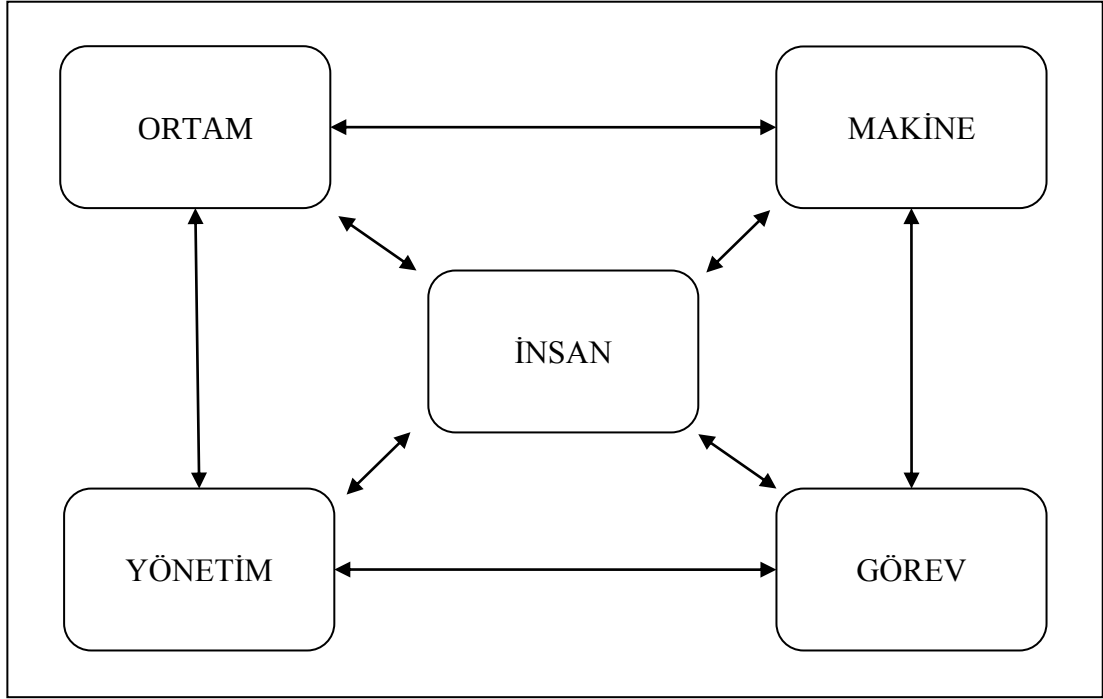
Bu çalışmada da kullanacak yöntem olan kaza etmenlerini beş ayrı bölümde inceleyen “5m” in açılımı:

- Man (İnsan),
- Machine (Makine),
- Management (Yönetim),
- Medium (Ortam),
- Mission (Görev) yöntemidir (Karakuş 2006).

Bu yöntem adım adım açıklamak gerekirse (Şekil 2.9.) :

- İnsan,

Kazaya neden olan etmenlerin ortasında insan bulunmaktadır ve insan bütün kaza etmenleri ile birebir ilişki içerisinde. Eğer kaza kırımların önlenmesi gibi bir amaç belirlendi ise insan etmeni hem kendi içinde, hem de diğer etmenler içinde



Şekil 2.9. 5M etmenleri ve arasındaki ilişki (Karakuş 2006’den değiştirilerek alınmıştır)

incelenmelidir. İnsan etmeni konusunda ayrıntılı bilgiler Bölüm 2’de 2.2. içerisinde verilmiştir.

- Yönetim,

Bütün organizasyonlarda olduğu gibi havacılık endüstrisinde de var olan hataların ve başarıların sorumlusu yönetimdir. Bu yüzden organizasyon içinde bulunan tüm elemanların, en başta insanın doğru bir şekilde çalışmasından yönetim sorumludur. Bu da insanları doğru bir şekilde bilinçlendirerek ve uygun koşullar sağlayarak yapılabilir. Bunun için kaliteli çalışma ortamı, sürekli denetleme ve periyodik eğitimler sağlanması önemlidir.

- Makine,

Teknolojideki büyük gelişmelere rağmen hala uçakların tasarım, imalat ve bakım faaliyetlerinden dolayı hala birçok kaza ve olay bu yüzden gerçekleşmektedir. Örnek vermek gerekirse; tasarım dehası olarak sestten hızlı gidebilen yolcu uçağı olarak üretilen “Concorde” uçaklarının kanatlarında meydana gelen önlenemez çatlaklardan dolayı uçuştan kaldırılması verilebilir. Hava araçlarının var olan bakım faaliyetlerini etkileyen ve ömrüne karar veren uçuş saatidir. Bir uçak ne kadar uçarsa o kadar ayrıntılı bakım faaliyetleri içerisinde bulunmalıdır. Bir uçak parçasının arızalanması için üç türlü durum mevcuttur. İlki uçağın tasarımdan dolayı arızalı olması ve ilk kullanımda bu arızanın ortaya çıkmasıdır. İkincisi parçanın beklenmedik bir zamanda rastgele bir arıza ile gayri faal olmasıdır. Sonuncusu, parçanın kullanımdan ötürü artık ömrünü doldurması ve arıza yapmasıdır. Tüm bu parça arıza olaylarının çözülmesi için Feyza Gürbüz ve ekibi (2009b) bir veri madenciliği çalışması ile parçaların arıza yapmadan önce tespit edilmesi konusunda çalışmalarda bulunmuşlardır.

- Ortam,

Ortam etmenini ikiye ayırarak incelemek gerekir. İlki doğal ortam, ikincisi insan yapısı ortam. Doğal ortam; havanın yapısı, arazi yapısı ve gerçekleşecek tüm doğa faaliyetleri (rüzgar, yağmur, bulut durumu gibi).İnsan yapısı ortam; havaalanı, pist yapısı gibi insan

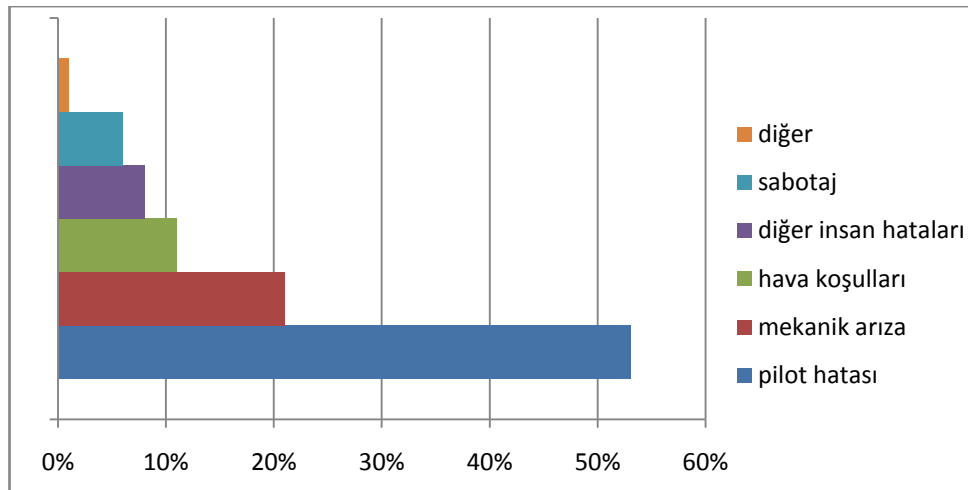
yapısı malzeme ve araçlar başlıkları altında incelenebilir. Ortam etmeni altında insan yapısı olanına müdahale imkanı varken, doğal ortam yapısına insanın doğrudan müdahalesi söz konusu bile olamaz. Fakat bundan kaçınarak kaza ve kırım azaltılmaya çalışabilir. Bunun için meteoroloji biliminden azami seviyede yararlanmak gerekir.

- Görev,

Her türlü görev aşamasında, birçok kaza ve olayın gerçekleşmesi potansiyel olarak mümkündür. Normal bir yolcu taşıma uçuşuna nazaran ağır yüklerle yapılan bir kargo nakliyesi, deniz üzeri yapılan uzun uçuşlar da kaza yapma riski daha yüksektir. Bu yüzden görev için yapılan uçuş planı, zaman ve saat önemlidir.

Bu çalışmada 5m kuralına ek olarak, kaza ve olaylara neden olduğunu düşündürecek etmenlerin insan hatası ile ilişkisini değerlendirmek için kaza ve olay sonuçlarının özellikleri de incelenmiştir. Bu bağlamda kaza sonucunda oluşan yaralı ve ölü sayısı, oluşan hasar ve kazaya oluşum nedenleri incelenmiştir.

İnsan, makine, bakım gibi etmenler arasında kazaların yarısının, özellikle ölümcül olanlarının nedeni insan hatalarıdır. Cenevre’de bulunan Uçak Kazası Kayıt Merkezi’nin raporuna göre 1950-2006 yılları arasındaki süreç incelendiğinde, 1843 uçak kazası olmuştur. Şekil 2.10.’da görüldüğü üzere kazaların nedenlerine göre dağılımın da insan hatasının toplamda %61 gibi yüksek bir orana sahip olduğu görülmektedir.



Şekil 2.10. 1950-2006 yılları arasında kazaların nedenlerine göre dağılımı¹

Daha önceki yıllarda insan etmeni başlığı altında genellikle pilotlar dikkate alınırdı fakat son yıllarda fark edildi ki; uçuşa dahil olan bütün insanlar incelenmelidir. Bu çalışmada, insan hatalarına neden olan alt etmenler incelenerek, uçak kazalarına olan etkileri araştırılacak ve azaltılması için önermelerde bulunulacaktır. Bunu da yaparken büyük hacimli verilerdeki bilgileri çıkarmadaki başarısından dolayı, veri madenciliği teknikleri kullanılacaktır. Veri madenciliği ile ilgili ayrıntılı bilgi Bölüm 3’te yer almaktadır. Uçak kazaları özellik olarak, çok az gerçekleşen fakat maddi ve manevi

¹<http://www.uted.org/dergi/2008/eylul/bakisacisi/nht.htm> (19.03.2010)

olarak büyük sonuçlar yaratan özel kazalardır. Unutulmamalıdır ki, aynı uçak içinde seyahat eden insanların, gerçekleşebilecek bir kazada kurtulma şansları, kara taşımacılığındaki kadar yüksek değildir. Bu özelliklerinden dolayı gerçekleşen uçak kazalarının neticeleri; yeryüzünde bir şekilde gerçekleşen büyük bir trafik kazasının ya da doğal afet gibi büyük zararlara neden olan bir doğa olayının yol açtığı can ve mal kaybı ile denklik gösterir. Daha açık belirtmek gerekirse; yolcu uçağında gerçekleşen olası bir kazanın önlenmesi, 100 adet insan kaybı ve 30 milyon dolar kadar maddi zararın da önlenmesi demektir. Yani kazaların ve olayların aslında %10 azaltılabilmesi bile, çok büyük olumlu sonuçlar doğurabilmektedir.

Örnek vermek gerekirse, 2009 yılında %11 oranında yolcu kapasitesini arttıran Türk Hava Yolları'nın¹ (THY) Hollanda'nın Amsterdam şehrinde gerçekleşen kazasında², yaşanan can ve mal kaybı tüm bunlara ek olarak da hesaplanamayacak derecede saygınlık kaybı yaşamasına yol açmıştır.

Belki de; 1000 uçuşta bir yaşanacak bu tür kazalar veya 100 uçuşta bir meydana gelebilecek ve kazaların habercisi olan olaylar, bunların yarattığı trajediler, yaşanan maddi ve manevi kayıpların hepsi eğer kazaların ve olayların verileri ayrıntılı şekilde incelenir ve kazaların etmenleri arasındaki ilişkiler ortaya çıkarılırsa, büyük ölçüde engellenebilir.

2.7. Konu İle İlgili Daha Önce Yapılan Çalışmalar

Daha önce veri madenciliğinin kazalar ve havacılık alanına uygulanmış örnekleri incelenirse:

- Ozgan 'Karayolu Araç Tipi ve Kaza Şekli ile Kaza Sonuçları Arasındaki İlişkilerin Analizi" adlı çalışmada:

2000 ila 2004 yılları arasında D100/11 karayolunda olan 783 kazayı incelenmiş ve araç tiplerine bağlı olarak kaza şekli ve kaza sonuçları arasında anlamlılık düzeyleri incelenerek tahmin modelleri çıkarmıştır. Bunu yaparken elde edilen verilere "SPSS istatistik programı ile çoklu lineer regresyon (doğrusal bağlanım), korelasyon ve varyans analizleri" uygulamıştır. Fakat bu çalışma tamamen istatistiksel yöntemlerle yapıldığı için çıkan sonuçlar yalnızca bir verinin diğer bir veri üzerine anlamlılık düzeyinin değerlendirilmesi ile açıklamıştır. Verilerin bir veya birden fazla veri ile olan ilişkileri sıralı olarak incelenememiştir (Ozgan 2008).

- Gürbüz ve arkadaşları 'Türkiye'de Bir Hava Yolu İşletmesine Ait Parça Söküm Raporlarına İlişkin Veri Madenciliği Uygulaması" adlı çalışmalarında;

Bir uçak firmasının parça söküm raporlarını veri madenciliği ile çözümleyerek uçak

¹ http://www.thy.com/tr-TR/corporate/news/press_room/partnership/passanger.aspx (10.04.2010)

² http://www.ubak.gov.tr/BLSM_WIYS/UBAK/tr/BELGELIK/guncel_haber/20090626_121048_204_1_1541.html (10.04.2010)

parçalarının arıza öncesi uyarı seviyelerinin kestirilmesine yönelik kural keşfi yapmışlardır. Bunu yaparken de Polyanalyst adlı veri madenciliği programı kullanılmış; sınıflandırma teknikleri altında lineer regresyon, yapay sinir ağı, bağımlılıkları bulma alt yöntemleri uygulanmıştır. Veriler arasındaki korelasyon lineer regresyon ile incelenerek var olan 18 adet parametreyi 16'ya indirgemişlerdir. Çalışmalarında elde ettikleri veriler ışığında kural bulma ve bağımlılıkları bulma yöntemleri en güvenilir sonuçları vermiş, var olan veri tabanlarında yer alan farklı birçok veri ile yapılabilecek veri madenciliği çalışmalarında da kullanılabileceği vurgulamışlardır (Gürbüz vd. 2009b).

- Han ve arkadaşları 'Kore Hava Kuvvetleri'ndeki Uçakların Sıralı Birliktelik Kuralı ile Arıza Modellerinin Tahmin Edilmesi' adlı çalışmalarında;

Modern savaşlarda hava kuvvetlerinin efektif olarak görev yapmasının en iyi yolunu önleyici bakım faaliyetlerinin uygulayarak elde edileceğini kabul etmişlerdir. Diğer veri madenciliği çalışmalarından farklı olarak birliktelik kuralının farklı şekli olan sıralı birliktelik kuralını uygulamışlardır. Sıralı birliktelik kuralında, normalinde farklı olarak işlem tabanlı yerine dizi tabanlı bir yapı benimsenmiş ve destek aralığı farklı bir şekilde hesaplanmıştır. Tasarladıkları bu modeli 2004 ve 2005 yılları arasında Kore Hava Kuvvetleri'ndeki dört adet uçak ilgili 16 arıza verisini içeren 38142 veri incelenmiş, bu veriler ile ilgili altı adet senaryo üretilmiş ve bu senaryolar altında dört mevsim için arıza modelleri değerlendirilmiştir. Böylece arıza modellerini ve ard arda gelecek arızaları tahmin ederek uçakların maksimum kullanımı amaçlanarak, uçak yedek parçalarının gelecekte kullanım talebini tahmin etmenin yolunu bulmuşlardır (Han vd. 2009).

- Gürbüz ve arkadaşları 'Ölümle Sonuçlanan Uçak Kazalarının Sınıflandırma Kuralı ile Keşfi' adlı çalışmalarında:

Amerikan Federal Havacılık kurumundan alınan uçak olayları verileri ile 25 adet kaza özelliği incelenmiş ve bu etmenler Rosetta adlı program kullanılarak aralarındaki bağımlılığa göre azaltım yapılmış, içlerinden ölüme neden olan sekiz adeti tespit edilerek Polyanalyst programı ile sınıflandırma tekniği uygulanmıştır. Bu çalışmada diğer çalışmalardan farklı olarak elde bulunan veriler öncelikle aralarındaki bağımlılıklara göre düzenlenmiş ve daha sonra veri madenciliği filtre edilen bu verilere uygulanmıştır.

Ayrıca çalışmanın ele aldığı konu özellikle ölüme neden olan etmenlerin bulunarak uçak olaylarının çözülmesidir (Gürbüz vd. 2009a).

- Letourneau ve arkadaşları "Uçak Parçalarının Değişiminin Veri Madenciliği ile Tahmin Edilmesi" adlı çalışmalarında:

Uçak üzerinde bulunan parçaların değişim zamanının tahmini yapılmaya çalışılmıştır. Uçak parçalarına sensör takılmış ve sensörlerden gelen veriler yer istasyonunda toplanmıştır. Amaç uçak parçalarının bozulmadan değişmesi gerektiğini tahmin ederek uçuş veri görüntüleme sistemine bütünleşmiş bir sistemle herhangi bir şekilde bir parçadan uyarı geldiğinde anında müdahale etmektir. Böylece kazalar önlenir, ertelemeleri ve bakım maliyetleri azaltılarak, bakım aktivitelerinin düzenlenmesi amaçlanmıştır. Çalışmada 34 adet Airbus A320 marka ticari uçağından ayda 1 MB veri alınarak bir veri seti oluşturulmuş, karar ağaçları, bayes, sinirsel ağlar, regresyon teknikleri ile çözümlemeler yapılmıştır (Letourneau vd. 1999).

- Bayam ve arkadaşları "Yaşlı Sürücü Kazalarının Veri Madenciliği Uygulaması ile Meta Analiz Edilmesi" adlı çalışmalarında:

Amerika'da gerçekleşen kazalarda yüksek risk taşıyan yaşlı sürücülerin demografik özellikleri araştırılmıştır ve bunu yaparken de yaşlı insanların kişisel özelliklerini

(yaşadıkları şehir, sağlık durumları v.b) içeren veriler SPSS Clementine programı kullanılarak veri madenciliği yöntemi ile incelenmiştir. Çalışmada uygulanan başlıca veri madenciliği yöntemleri, sınıflandırma kuralları altında yer alan karar ağaçları ve yapay sinir ağı çözümleridir. Çalışmadaki amaç yaşlı sürücülerin kazalara karışmasını azami seviyede azaltmaktır. Çalışmanın yapılış aşamasında diğer çalışmalardan farklı olarak hazır veriler değil de sürücülere anket uygulanarak ve anketlerden elde edilen veriler SPSS programı ile işlenerek veri madenciliği uygulaması için hazır hale getirilmiştir (Bayam vd. 2005).

- Marukatat “Trafik Kaza Verilerinin Yapısal Tabanlı Birliktelik Kuralı ile Kural Bulma Çerçevesinde İncelenmesi” adlı çalışmasında;

Bankok’un kuzeyinde gerçekleşen trafik kazalarını birliktelik kuralı uygulayarak incelemiştir. Daha önceki birliktelik kuralları ile aynı fakat yeni özellikler sunmuştur. Bu çalışmanın diğerlerinden farkı; uygulanan çözümleme sonucunda elde edilen kurallar dört ana kategoride zayıf, az verimli, verimli ve çok verimli sınıflandırılmış olmasıdır. Böylece klasik birliktelik kuralı değerlendirmelerinden ayrılmıştır (Marukatat 2006).

- Nazeri ve arkadaşları “Veri Madenciliği Yaklaşımları İle Uçak Kazaları Ve Olayları Arasındaki İlişkinin Analiz Edilmesi” adlı çalışmalarında;

Temel olarak veri madenciliği üzerinde durulmasa da havacılık alanı için çok önemli olan iki konu; uçak olayları ve uçak kazaları arasındaki bağlantı araştırılmıştır. Bunu yaparken de İsviçre peyniri ve Heinrich piramidi modelleri uygulanmıştır. İsviçre peyniri modelinde; her olay bir peynir diliminin deliğinden geçer ve her dilim kazayı önleyen bir sistemdir. Sonuçta kaza sürecinde her dilim tek başına bir kazaya neden olmaz fakat bütün dilimler üst üste gelirse işte o zaman kaza oluşur. Bir diğer yöntem olan Heinrich piramidinde en yüksek risk düşük frekans en üsttedir. Sonuçta piramidin altına doğru ilerledikçe risk azalmaya başlar fakat frekans değerleri artar. Heinrich piramidinde amaç kazalar ve olaylar arası sayısal ilişkiyi belirlemektir.

Nazeri ve arkadaşları bu çalışmasında temel olarak kazaya ve olaylara neden olan etmenler arası korelasyon kurar bunu da daha çok İsviçre peyniri modeli ile gerçekleştirir.

Çalışmalarının sonucunda uçak olaylarının eksik ve yanılgılı olabileceğini, fakat uçak kazalarının tam ve tarafsız olduğunu vurgulamışlardır. Buna da neden olarak uçak olaylarının bakım ve kabin ekibi tarafından doldurulan formlarla tespit edilip kayıt altına alınması, kazaların ise tarafsız uluslar arası bir kurul tarafından her yönden incelenmesi olduğu sonucuna varmışlardır.

Nazeri ve arkadaşlarının ulaştığı nokta ışığında çıkan önemli sonuçlardan biri ve bu çalışmaya da yön veren durum; normal koşullarda kötü hava koşullarının kendi başına uçak olayı etmeni olarak adlandırılmasıdır fakat bu etmen pilot hatası ile birleştirilince artık bu olay kaza etmeni adını alır (Nazeri vd. 2008).

- Harris ve arkadaşları “Veri Madenciliği Yaklaşımı İle Havacılık Güvenliğindeki Son Yenilikler” adlı çalışmalarında;

Asıl gaye olarak diğer bütün çalışmalarda da olduğu gibi kazaları önlemek için veri madenciliği çözümlemesini en etkin şekilde Quinlan C4.5 karar ağacı algoritması ile kullanmaktır. Bu çalışmayı diğerlerinden farklı yapan çalışma sonucunda elde edilen sonuçların sıradan bilinen değil de bir kaza kırım inceleme uzmanını bile şaşırtacak derecede sonuçlar elde etmek için çalışılmış olmasıdır. Bu çalışmada en başta

geleneksel veri madenciliği yöntemi ile geleneksel sonuçların ortaya konduğu ispatlanmış ve daha sonra kendi ürettikleri Smither isimindeki algoritmayı Quinlan C4.5 algoritmasının yerine koyarak yeni ve ilginç sonuçlar elde etmeyi hedeflemişlerdir (Harris vd. 1998).

- Saraee ve arkadaşları, “Ölümle Sonuçlanan Çocuk Kazalarını İçeren Medikal Verilere Veri Madenciliği Analizi Uygulanması” adlı çalışmalarında:

15 yaş altı çocuklarda ölüme neden olan travma olayları incelemiş, inceleme esnasında SPSS Clementine veri madenciliği programı ile sınıflandırma analizi altında karar ağaçları ve naif bayes yöntemleri karşılaştırılmış ve karar ağaçlarının üstün olduğu ve daha doğru sonuçlar çıkardığını ispatlamışlardır (Saraee vd. 2008).

- Nazeri ve arkadaşı, “Şiddetli Hava Koşullarının Hava Sahası Sistem Performansına Etkilerini Anlamak İçin Havacılık Verilerinin Analiz Edilmesi” adlı çalışmalarında:

Hava durumu ve trafik verilerini incelenmişlerdir. Amaç hava durumunun özellikle kötü hava koşullarının uçuş trafiğine nedenli etkilediğini, rötaların ve uçak kazalarını nasıl en aza indirilebileceğini bulmaya çalışmışlardır. Bunu yaparken de sınıflandırma ve kümeleme analizleri uygulamışlardır. Sonuç olarak da bir uçuş aracının karşılaşılabileceği en kötü hava koşulunun, kalktığı veya indiği meydanın hava koşulları değil de, uçuş esnasında karşılaşılabileceği hava olduğu sonucuna varmışlardır (Nazeri ve Zhang 2002).

- Brown ve arkadaşları, “Kaza Datalarının Lojistik Regresyon ve Karar Ağaçları ile Analiz Edilerek İnsan Hataları Dinamiklerinin Araştırılması” adlı çalışmalarında:

Son yıllarda gerçekleşen tren kazalarında ve yolculuklarında tüm inisiyatifin operatörlerin kontrolüne verildiğini ve insan etmeninin en ön plana çıktığından dolayı tren yolculuklarında kaza risklerinin arttığını vurgulamışlardır. Amaçları kazalarda insan etmeninin önemini araştırmaktır ve yine bunu sınıflandırma analizi ile yapmışlardır (Brown vd. 2002).

- Pande ve arkadaşı, “Geniş Yetki Alanındaki Kaza Verilerine Pazar Sepeti Analizinin Uygulanması ve Potansiyel Karar Destek Aracı” adlı çalışmalarında:

Birliktelik kuralının en çok kullanılan modeli olan pazar sepeti analizi uygulanmıştır. Çalışmada Amerika/Florida eyaletindeki kaza verileri üzerine çalışılmıştır. Bu çalışmanın farkı pazar sepeti analizini yaparken apriori algoritmasında kısmi değişiklikler yapılmasıdır. Algoritma uygulanırken minimum destek değeri düşürülmüş ve bu da eşik değeri olarak atanmıştır. Böylece düşük destek değerinin üzerindeki her değer bir kural olarak ortaya çıkmış, yüksek sayıda kural elde etmişlerdir. Elde edilen yüksek sayıdaki kural ile daha sağlıklı çıkarımlar elde etmeye çalışmışlardır (Pande ve Abdel-Aty 2007).

Yapılan tüm çalışmaların ışığında bu çalışmada; uçak kazalarında daha önce başlı başına tek bir etmen olarak incelenmeyen insan etmeni ele alınacaktır. Bu çalışmayı da gerçekleştirilirken veri madenciliği yöntemlerinden olan sınıflandırma ve birliktelik kuralı aynı veri dizisine uygulanarak hangisinin daha doğru sonuçlara ulaştırdığı irdelenecek ve güvenilirliği test edilecektir.

3. VERİ MADENCİLİĞİ

3.1. Giriş

Veri madenciliği büyük veri kümeleri içerisinde daha önce bilinmeyen, geçerli desenleri ve ilişkileri keşfetmek için çok yönlü veri çözümleme araçları içerir. Bu araçlar istatistiksel modeller, matematiksel algoritmalar ve makine öğrenme metotları (karar ağaçları ve sinirsel ağlarda deneyim ile performansı otomatik olarak arttıran algoritmalar) içerir. Sonuç olarak, veri madenciliği, verileri toplama ve yönetiminden daha fazlasını içerir, aynı zamanda çözümleme ve kestirim yapma olanağı tanır (Seifert 2006).

Veri madenciliği ile ilgili yapılan tanımlara göz atmak gerekirse:

- Veri madenciliği çoğu kez büyük ve gözlemsel veri kümelerinde daha önce şüphe duyulmamış ilişkileri bulan ve veriyi bir roman gibi özetleyerek, veri kullanıcısına anlaşılabilir/ kullanışlı bir bilgi şeklinde sunan çözümleme şeklidir (Hand vd. 2001).
- Veri madenciliğini Mitra ve Acharya (2003); veri tabanlarında bilgi keşfi ile eşanlamlı olarak, geçerli, yeni, potansiyel olarak faydalı ve en sonunda anlaşılabilir desenleri verilerden elde edebilecek bir süreç olarak tanımlamaktadırlar. Ayrıca toplam sürecin alçak-düzey veriyi yüksek düzeyde bilgiye dönüştürmeyi sağladığını belirtmişlerdir.
- Veri madenciliği önceden bilinmeyen ve potansiyel olarak yararlı bilgilerin önemsiz gibi gözüken saklı verilerden çıkarılmasıdır (Frawley vd. 1992).
- Veri madenciliği istatistik, veritabanı teknolojisi, örüntü tanıma, makine öğrenme ile etkileşimli yeni bir disiplin ve geniş veritabanlarında önceden kestirilemeyen ilişkilerin çözümlemesi olarak tanımlanmıştır (Akbulut 2006).

Veri madenciliği sayısal, metinsel ve çoklu ortam şeklinde temsil edilen veriler üzerine uygulanabilir. Veri madenciliği uygulamaları çeşitli öğeleri kullanarak verileri inceler.

Veri madenciliği öğelerini kısaca açıklamak ve örneklendirmek gerekirse:

- Birliktelik (Bir olayın diğer olaya bağlı olduğu desenlerdir)
Bir kalem ile bir kağıdı satın alma,
- Dizi analizleri (Bir olayın başka bir olaya yol açtığı desenler)
Bir çocuğun doğması ile çocuk bezinin satın alınması,
- Sınıflandırma (Yeni desenlerin tanımlanması)
Boru kaplama bandı ile plastik kaplama arasındaki rastlantı,
- Kümeleme (Önceden bilinmeyen olguların bulunması ve görsel olarak belgelerin gruplanması)
Coğrafi konum ile marka tercihleri,

- Öngörü (Desenleri keşfederek, gelecekteki faaliyetlerle ilgili akılcı kestirimlerde bulunmak)
Atletik kulübe üye olan insanların egzersiz dersi alma ihtimali gibi.

Bir uygulama olarak, diğer veri çözümleme uygulamalarıyla karşılaştırıldığında, dereceden daha çok tür olarak farklılık gösterir. Örneğin yapılandırılmış sorgu (birçok ticari veritabanında kullanılır.) veya istatistiksel analiz yazılımı gibi. Çok basit analitik araçlar doğrulama temelli yaklaşımlar kullanır, kullanıcı bir hipotez geliştirir, daha sonra hipotezi kanıtlamak veya çürütmek için veriler test edilir. Örneğin, bir çekiç alan müşteri aynı zamanda bir kutu çivi satın alır şeklinde hipotez geliştirebilir. Bu yaklaşımın etkinliği kullanıcının çeşitli hipotezler geliştirme yaratıcılığının yanı sıra kullanılan yazılımın yapısı ile sınırlıdır.

Buna karşılık, veri madenciliği algoritmaları ile çok boyutlu veri ilişkilerini aynı zamanda incelemek, onların bir kez veya sık sık yinelandığını tanımlamak için keşif yaklaşımını kullanır. Örneğin, bir yapı malzemeleri mağazası müşterilerini; ev sahibi olmalarına, kullandıkları otomobile, yaş, meslek, gelir düzeyi, evleri ile mağaza arasındaki mesafeye göre karşılaştırabilir. Bu karmaşık yeteneklerinin bir sonucu olarak, başarılı bir veri madenciliği araştırması için iki ön madde önemlidir:

1. Problemin çözülmesi için net bir formülasyon,
2. İlgili verilere erişim.

Veri madenciliğinin bu kavramlaştırma yansıtmasına rağmen, bazı gözlemciler veri madenciliğinin büyük bir süreç olarak bilinen veritabanlarında bilgi keşfinde yalnızca bir basamak olduğuna dikkat çekmektedirler (Frawley vd. 1992).

Teknolojik gelişmelerdeki ve iş süreçlerindeki bir dizi gelişme, veri madenciliğine hem kamu hem de özel sektörde ilginin büyümesine katkıda bulunmuştur. Veri tabanlarının daha derinlemesine incelenebilmesine zemin hazırlayan gelişmeler; bilgisayar ağlarındaki büyüme ile veritabanlarının birbirine bağlanabilmesi, sinirsel ağlar ve ileri algoritmalar gibi arttırılmış arama ile ilgili tekniklerin geliştirilmesi, istemci/sunucu bilgisayar modellerinin yayılması, kullanıcılara bilgisayarlarından merkezi veri kaynaklarından erişim sağlanması ve bambaşka kaynaklardaki verilerin aranabilir tek bir kaynağa birleştirilebilme yeteneğinin artması bu değişikliklerden bazılarıdır (Akbulut 2006).

Bu gelişmiş veri yönetim araçlarına ek olarak, bilginin kullanılabilirliğinin artması ve depolama maliyetlerinin azalması önemli rol oynamıştır. Geçtiğimiz birkaç yıl içerisinde toplanan ve depolanan bilgi hacminde hızlı bir artış olmuştur, bazı gözlemciler, dünyadaki veri miktarının miktar olarak her sene yaklaşık olarak iki katına çıktığını belirtmektedirler (Han ve Kamber 2000). Aynı zamanda, megabayt başına 1.5 TL olan veri depolama maliyetleri önemli ölçüde düşerek megabayt başına 0.15 kuruşa gerilemiştir. Benzer şekilde, bilgisayar işlem gücü her 18-24 ayda bir artmakla beraber, işlem gücü maliyetleri de azalmaya devam etmektedir.

Veri madenciliği hem kamu hem de özel sektör için ortak bir konu olmuştur. Organizasyonlar veri madenciliğini müşteri bilgilerini araştırmak için bir araç olarak kullanmaktadırlar, dolandırıcılığı ve tüketimi azaltmada, tıbbi araştırmalarda bir yardımcı olarak kullanmaktadırlar. Ancak veri madenciliğinin çoğalmasıyla birlikte uygulama ve denetim sorunlarında artışlar da olmuştur. Bu sorunlar; çözümlenen verinin kalitesi, veri tabanlarının işlenebilirliği ve kuruluşlar arasında yazılım ve potansiyel gizlilik ihlalleri hakkındadır. Ayrıca, veri madenciliğinin limitleri konusunda, ülke güvenliğine yapılan girişimlerini vurgulamak için devlet organlarının üzerinde durduğu çalışmalar gibi gözden kaçanlar yüzünden endişeler oluşmuştur.

3.1.1. Veri madenciliğinin gelişimi

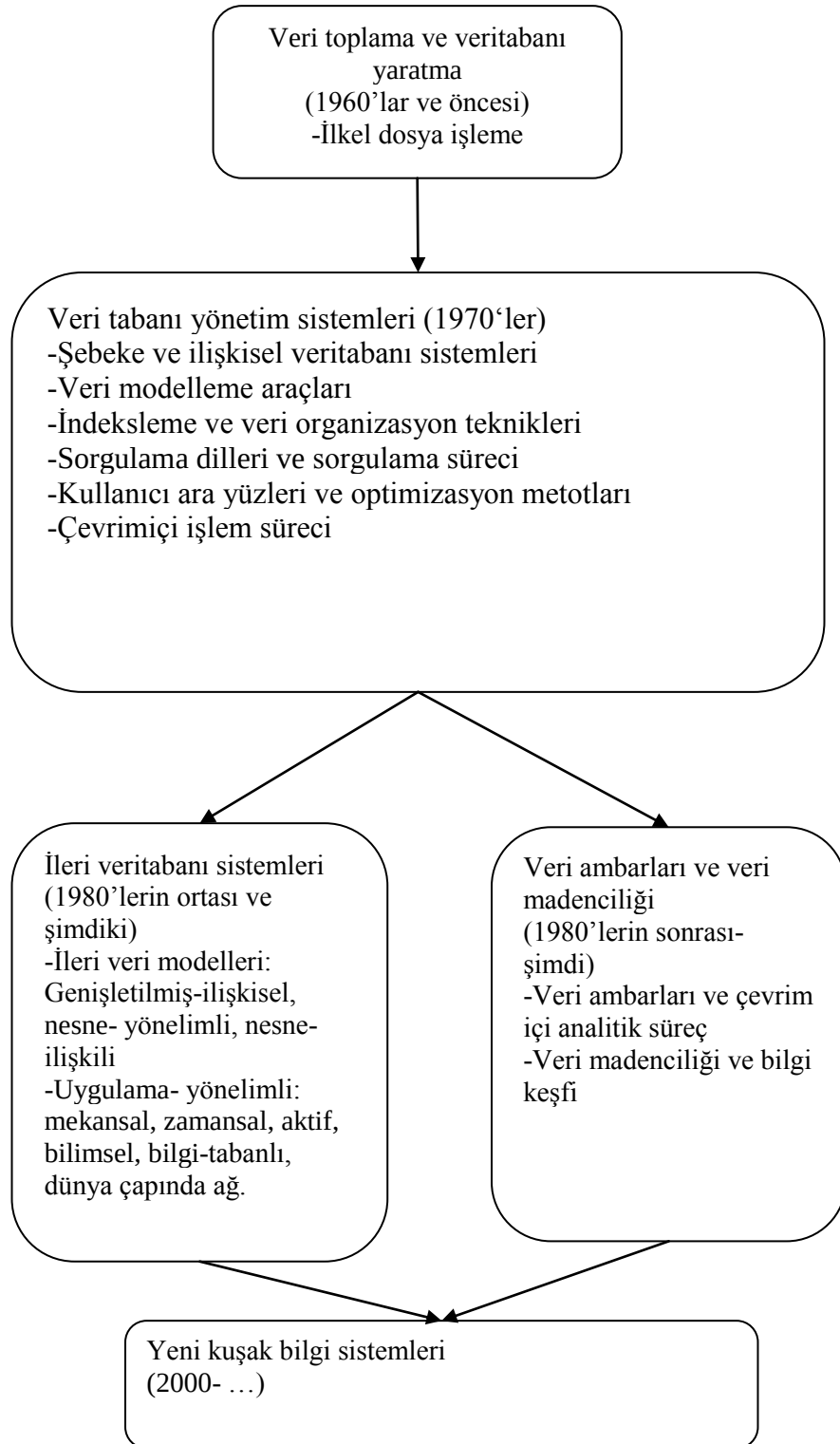
Veri madenciliğinin bilgi sektörünün büyük ölçüde dikkatini çekmesindeki önemli neden; büyük miktarlardaki veriyi tamamen kullanabilme ve bu tür verilerin yararlı bilgi ve tecrübeye dönüştürmedeki acil gereksinimdir. Kazanılan bilgi ve tecrübe; işletme yönetimi, üretim kontrol, pazar analizi, güvenlik, mühendislik tasarımı ve bilim araştırmaları gibi birçok uygulama için önemlidir.

Veri madenciliği bilgi teknolojisinin doğal evriminin bir sonucu olarak ortaya çıkmıştır. Veri tabanı alanında bulunan işlevler gerçekleşirken evrimsel bir yol izlemiştir. Veri toplama ve veritabanı oluşturma, veri yönetimi (veri depolama ve erişimi, veritabanı işleme süreci) ve veri analizi ve anlaşılması (veri ambarları ve veri madenciliği). Örneğin, veri toplama ve veri tabanı oluşturma mekanizmalarında gerçekleşen erken gelişme, daha sonra gerçekleşen veri depolama ve erişim, sorgulama ve işleme sürecinde gelişen etkili mekanizmalar için bir ön koşul olarak sunulmuştur.

Veritabanı sistemleri; sorgulama ve işleme sürecini yaygın bir uygulama olarak kısaltmaktadır. Veri çözümlemesini ve anlamlandırılmasını süreç olarak kısaltılmak için ikinci hedeftir.

1960'lı yıllarda veritabanı ve bilgi teknolojisi; ilkel dosya işleme sistemlerden ileri ve güçlü veritabanları sistemlerine doğru sistematik olarak gelişmektedir (Hand vd. 2001). Veritabanı sistemleri içinde araştırma ve geliştirme, 1970'li yıllarda ilişkisel veritabanı sistemleri (verinin saklandığı ilişkisel tablo yapıları), veri madenciliği araçları, dizin oluşturma ve veri organizasyon teknikleri gelişme göstermiştir. Ayrıca, kullanıcılar pratik ve esnek sorgulama dillerinde veri erişimi, sorgulama işlemi ve kullanıcı ara yüzleri elde etmişlerdir. Bir sorgulamanın salt okunur işlem olarak ele alındığı, etkili depolama, geri alma ve büyük miktarlarda veri yönetimi için önemli araç olan çevrimiçi işlem sürecindeki (OLTP) etkin metotların, geniş olarak kabul edilmiş ve evrimleşmiş ilişkisel teknolojiye büyük ölçüde katkısı vardır (Han ve Kamber 2000).

Çizelge 3.1. Veri madenciliği tarihçesi (Han ve Kamber 2000’den değiştirilerek alınmıştır)



1980'lerin ortasından bu yana veritabanı teknolojisi, ilişkisel teknolojinin popülerleşmesi ile yeni ve güçlü veritabanı sistemlerindeki araştırma- geliştirme faaliyetlerinin artışı ile tanımlanmıştır. Dağılım, çeşitlilik ve verinin paylaşımı ile ilgili

sorunlar kapsamlı olarak ele alınmıştır. Heterojen veri tabanı sistemleri ve internet tabanlı küresel bilgi sistemleri gibi (World Wide Web; WWW) ortaya çıkmış ve bilgi alanında önemli bir rol oynamaya başlamıştır.

Son yıllar içerisinde bilgisayar donanım teknolojilerindeki şaşırtıcı ve sürekli gelişim güçlü, satın alınabilir, büyük bilgisayar kaynakları, veri toplama araçlarına ve medya depolarının ortaya çıkmasına önderlik etmiştir. Bu teknoloji veritabanı ve bilgi endüstrisi için büyük bir destek olmaya başlamıştır. Ayrıca yüklü veritabanları, bilgi ambarları için işlem yönetimi, bilgi alımı ve veri çözümlemesi yapmaktadır.

Veriler şimdi farklı veritabanlarında saklanabilmektedir. Son zamanlarda ortaya çıkmış bir veritabanı mimarisi ile veri ambarı ve heterojen veri kaynakları deposu altında depolanmış veriyi, birleştirmekte ve sonuç üretmeyi kolaylaştırmaktadır. Veri ambarı teknolojisi; veri temizlemeyi, veri bütünleştirme ve çevrim içi analitik süreci içerir. Bunun yanında özetleme, pekiştirme ve kümeleme gibi çözümleme teknikleri ile farklı açılardan bilgiyi görme yeteneği de vardır. Her ne kadar çevrimiçi analitik süreç araçları çok boyutlu analiz ve karar vermeyi desteklese de, yardımcı araçlar(sınıflandırma, kümeleme ve zaman içindeki veri değişimi gibi.) derinlemesine çözümleme için gereklidir.

Veri bolluğu, güçlü veri çözümleme araçları ile birleştiğinde, veri bakımından zengin fakat bilgi bakımından fakir bir durum ortaya çıkar. Bütün sorun; verilerin amaca yönelik olarak kullanılabilirliği olmasından/ olmamasından kaynaklanmaktadır.

3.2. Bir Veri Madenciliği Projesinde Başarı İçin Genel Koşullar

Veri madenciliği projesinin başarısı için genel koşullar şunlardır:

A) Önemli bir kazanç beklenir:

- Genellikle daha önce denenmemiş analitik tekniklerin olduğu, sonuçları en kolay kazanılan hedef belirlenir.
- Geliştirilmiş kaldıraç etkisi yaratabilir. Öyle ki hayati bir süreçte artan bir gelişme güçlü sonuç etkisi yaratabilir. Örneğin uçak kazalarındaki kaybın % 10 dan % 9,8'e düşmesi milyonlarca Türk liralık maliyet farkı yaratabilir.

B) Yetenekli bir ekip veri madenciliği için gereklidir: Çok küçük projeler dışında, bir kişi tüm faaliyetlerde başarılı olamaz. Bir kişinin; veri çıkarımı, veri bütünleştirme, analitik modelleme, rapor oluşturma ve sunum için yeteri kadar zamanı olamaz.

- C) Veri uyanıklığı yani biriken bilgi akışının yakalanması ve toplanması (örneğin uçak kazalarını engellemek için uygulanan bir güvenlik modelinden sonra yeni verilerin toplanması)
- D) Zaman yani öğrenme; birçok model üretim yinelemeleri ve çevrimden sonra gerçekleşebilir. Örneğin Dr. Ferdinand Porsche'nin kurumsal anlayışı "Yarışmak üretmeyi artırır." öğretisidir. Bugün, Porsche küresel olarak en karlı otomobil üreticilerinden biridir.

Aynı şekilde, veri madenciliği modelleri gerçekliğe karşı yarış içerisinde olmalıdır. İlk başta üretilen veri madenciliği modelleri daha kısır ve ilkeldi; sorgulama dillerinin de gelişmesiyle başarı oranı artırılmış ve günümüz veri madenciliği modelleri ortaya çıkmıştır. Model üretim yinelemeleri, başarının artmasını sağlamıştır. Bu türden, her veri madenciliği uygulaması prensip olarak aynı ortak yöntemi takip eder.

3.3. Veri Madenciliği Projesini Başarıya Ulaştırmak için Dikkat Edilmesi Gereken Hususlar

Hand vd.(2001) veri tabanlarında örüntülerin keşfi için kullanılan veri madenciliği araçları için bazı uyarıları şöyle özetlemiştir:

- Veri kalitesi:
Düşük veri kalitesi veri madenciliği metodlarının açıkça ortaya çıkmasını engeller. Düşük veri kalitesi zayıf modeller ürettirecektir. Yine de zayıf veriler yüksek kestirim yeteneğine sahip modelleri destekleyebilir fakat bu modeller hayali olabilir.
- Müdahaleler:
Veri madenciliği modelinden elde edilen sonuçlar ile üretilen çözümler çok kısa zaman da geçerliliğini yitirebilir, bunu engellemek için bazı değişiklikler yapılabilir. Örneğin; gelişmekte olan sahtekârlık algılama modelleri bazı etkili kısa vadede önleyici tedbirlere neden olabilir. Fakat bundan hemen sonra, dolandırıcılar operasyonlarında bu müdahaleleri önlemek için yeni davranışlar geliştirebilirler.
- Ayrılabilirlik:
Çoğu zaman, veri gruplarından elde edilecek ilgi çekici bilgiyi sıradan bilgilerden ayırmak zordur. Veri gruplarında birçok desen olabilir, fakat verilen problemi çözmek için veri madencisini yalnızca birkaç tanesi ilgilendirebilir. Hedef değişkenlerin tanımlanması, hangi algoritmanın kullanılacağını belirlemek için en önemli etmenlerden birisidir. Örneğin, hedef değişkenini tanımlamak için son iki fatura döneminde müşteri etkinliğinde %70 düşüş en iyi yol olabilir.
- Açıklık:
Veri kümesinden keşfedilen bazı desenler kullanışlı olmayabilir çünkü bu desenler veri madenciliği çözümlemesi olmadan da oldukça açık ve nettirler. Örneğin, bir veri setinde evli erkekler ile evli kadınların sayısı eşittir. Ya da

yumurtalık kanserinin en başta kadınlarda oluşması ve yine çek dolandırıcılığının daha çok çek hesabı olan müşterilerde meydana gelmesi gibi.

- Durağan olmaması:
Veri madenciliği çözümlemesinde kullanılan veri grubuna eklemeler yapılabilir veya tamamen silinebilir. Örneğin, üzerinde çalışmalar yapılan bir geyik nüfusu, oluşabilecek doğa seleksiyonlardan dolayı azalabilir veya tamamen yok olabilir.

3.4. Veri Madenciliğinde Karşılaşılan Önemli Sorunlar

Veri madenciliğinin işlerliğini sürdürebilmesi için gereken ham verinin eksiksiz, hatasız ve sorunsuz olması gerekir. Verinin uygun olmamasından başka veri madenciliğini etkileyen birçok sorun daha vardır. Sınıflandırmak gerekirse başlıca sorunlar aşağıdaki gibidir¹.

- Değişik türdeki bilgilerin farklı veritabanlarından elde edilmesi:
Verilerin çeşitli girdi kaynaklarından birleştirilmesi gerekli olabilir. Bu girdi kaynakları veri ambarları/ veri pazarları, Excel tabloları, metin belgeleri ve görüntü verileridir. Bu bütünleştirme oldukça karmaşık ve zaman alıcı olabilir.
- Çeşitli seviyelerdeki bilginin interaktif veri madenciliği ile etkileşimi:
Hesap düzeyindeki veri, birey düzeyindeki veri ile birleşik olmalıdır ve farklı zaman çizgisi içeren (günlük, aylık v.b.) veri ile koordineli olmalıdır. Bu sorun her türlü giriş verisinin birbiriyle tutarlı yapmak için dikkatli şekilde dönüştürülmesini gerektirir.
- Arka plan bilgilerinin birleştirilmesi:
En güçlü kestirim değişkenlerinin bir kısmı, ortak veritabanının dışından toplanmıştır. Bu veriler, demografik, geçmiş bilgi içeren ve üçüncü şahıs verileri içeriyor olabilir. Bu dış verilerin iç verilerle bütünleştirilmesi çok zor ve belirsiz olabilir. Hatalı eşleşme birçok durumda meydana gelebilir. Ayrıca bu süreç çok zaman alabilir.
- Veri madenciliği sonucunda verilerin sunumu ve görselleştirme:
Son derece teknik verilerin, teknik olmayan yöneticilere sunumu çok zor olabilir. Verilerin grafik ve görsel olarak sonuçları, analitik becerileri sayısaldan çok grafiksel olan idareciler için çok değerli olabilir.
- Gürültülü ve eksik veri ile başa çıkma:
Verilerin çoğu öğeleri boştur. Örnek olarak müşteri kaydı veya hesap kaydı genellikle boştur. Veri madenciliğinde en zorlu görevlerden biri bu boşlukları sezgisel değerlerle doldurmaktır. Orada olmayan verilere ek olarak, kayıtlarda gösterilen bazı veriler rastgeledir ve gürültü oluştururlar. Farklı veri madenciliği algoritmalarının veri eksiklerine ve gürültüye farklı duyarlılıkları vardır.

¹http://tunedin.entrepreneur.com/tradejournals/article/202562262_5.html (05.05.2010)

- Desen değerlendirme- ilginç olmama problemi:
Veri kümesi üzerinde birçok desen olabilir. Veri madenciliği için zor olan bu desenlerden ilginç olanlarını ve eldeki veri madenciliği sorununu çözmek için yararlı olanını ayırt etmektir. Önerilen desenlerden ilgi çekmeyenleri, kullanıcının ilgi sahasına göre çeşitli ölçütlerle seçilir ve derecelendirilir. İlgi çekmeyen desenleri belirlemek için iyi ölçütler belirlemek bu değişkenleri vurgular ve bunun gereksiz değişkenlerin silinmesine ve modele çok önemli katkısı vardır. Bu etkinlik, zaman kazancı sağlar.
- Veri madenciliği algoritmalarının verimliliği ve ölçeklenebilirliği:
Bir veri madenciliği algoritmasının verimliliği onun kestirim gücü ve bir model yaratmak için harcadığı zaman kavramlarına göre ölçülür. Ölçeklenebilirlik sorunları bir model veya algoritmanın küçük veri kümesi üzerinde kurulduktan sonra büyük veri gurubuna uygulanması ile ortaya çıkar. İyi veri madenciliği algoritma ve modelleri doğrusal olarak ölçeklenebilir; süreç içerisinde harcanan zaman veri setinin karesel olarak artmasına rağmen geometrik olarak artar. Bu da önemli bir etmendir.
- Paralel, dağıtılmış ve artan veri madenciliği algoritmaları:
Büyük veri madenciliği problemleri, bölerek ve yöneterek çok işlemcili paralel bilgisayarlarda çok daha verimli şekilde çözülebilir. Büyük veri kümelerini analiz etmek için başka bir strateji, işlemi birden çok bilgisayarla yapmak için dağıtarak yapmak ve sonuçları bütün çıktıları bütünleştirerek oluşturmaktır. Son olarak bazı veri madenciliği problemleri (örneğin elektrik şebekesi problemleri) artan algoritmalar kullanarak veya sabit büyük veri yığını üzerinde değil de sürekli kendini yeni verilerle yenileyen veri kümesi üzerinde çalışarak çözülür.
- Verilerin ilişkisel ve karmaşık türlerinin kullanımı:
Veriler karmaşık tabloların bir araya gelmesiyle oluşabilir. Birçok girdi verisi, ilişkisel tablo halindeki veri tabanlarından gelebilir (ortak anahtarla birbirine bağlanmış standartlaştırılmış tablolardan oluşan sistem). Diğer girdi verileri karmaşık çok boyutlu veritabanlarından gelebilir (yıldız-şemaların ayrıntıları). Veri madenciliği süreci her ikisini de içine alacak şekilde yeterince esnek olmalıdır.
- Küresel ve heterojen bilgi sistemlerinde veri madenciliği:
Veri madenciliği araçları, çok farklı veritabanı yapılarında veri girişi işleme yeteneğine sahiptir. Grafiksel kullanıcı ara yüzü araçlar da, birden fazla düğüm, çok farklı veri yapılarına veri girişi için yapılandırılabilir.

3.5. Veri Madenciliği Kullanım Alanları.

Saklı duran yararlı bilginin geniş hacimli verilerden keşfetme yeteneği ve o bilgi üzerinde etkili olabilmek bugünün rekabetçi dünyasında giderek önem kazanmaktadır. Çeşitli alanlardaki şirketler veri madenciliği ile geniş veri tabanlarını değerli, detaylı işlem bilgileri elde ederek rekabet avantajı elde etmek için kullanılmaktadırlar. Veri madenciliği metotlarının kullanıldığı çok çeşitli sahalarda ve kullanım amaçları ana başlıklar altında aşağıdaki gibi özetlenebilir: (Rygielski 2002)

Aşağıda verilen her dört uygulama, veri madenciliğinin iki aktivitesini gerçekleştirmektedir. Keşfetme ve tahmin etme. Keşfetme süreci, örneklerde açıkça sözü edilmemiş (perakendecilik tanımı hariç), müşteri bölümlmek için kullanılmıştır. Bu işlem de koşullu mantık, birliktelik ve benzerlik analizi, yönelmeler ve varyasyonlar ile yapılır. Aşağıda tanımlanan bütün uygulama kategorileri bir tür tahmin edici modellemedir.

3.5.1. Perakendecilik

Mağaza, markalı kredi kartı ve nokta satışı sistemler yoluyla, perakendeciler her alışveriş işleminin ayrıntılı kayıtlarını tutabilmektedir. Bu onların çeşitli müşteri kategorilerini daha iyi anlamalarını sağlar. Bazı perakendecilik uygulamalarına örnek vermek gerekirse:

- Sepet analizi yapmak: Benzerlik analizi olarak da bilinen, sepet analizi; hangi malzemelerin müşteriler tarafında birlikte satın alma eğiliminde olduğunu ortaya koymaktadır. Bu bilgi stoklamayı, mağaza düzeni ve özendirme stratejilerini geliştirebilir.
- Satış tahmini: Zaman tabanlı modelleri incelemek perakendecilerin stok kararları almalarına yardımcı olur. Eğer bir müşteri bir malzemeyi bugün satın alırsa, bu malzemeyi tamamlayıcı malzemeyi ne zaman hangi olasılıkla satın alır sorusuna yanıt bulabilir.
- Veritabanı pazarlama: Perakendeciler kesin davranışları olan müşterilerinin profillerini geliştirebilirler. Örnek olarak hep aynı marka giysiyi satın alan veya büyük satışlara her zaman katılanlar gibi. Bu bilgiler, etkin-maliyet özendirici tanıtımlarına odaklanmak için kullanılabilir.
- Stok planlama ve tahsisi: Perakendeciler yeni mağaza açtıkları zaman, stok planlama ve tahsisini; benzer demografik özelliklere sahip yerlerdeki mağazalarındaki desenleri inceleyerek geliştirebilirler. Perakendeciler ayrıca veri madenciliğini kullanarak, belirli bir mağaza için uygun düzenlemeyi belirlemek için kullanabilirler.

3.5.2. Bankacılık

Bankalar çeşitli uygulamalar için bilgi keşfinden yararlanabilirler (Akbulut 2006). Örneklerinden söz etmek gerekirse:

- Kart pazarlama: Müşteri kategorilerini tanımlayarak, kart satıcıları daha etkili satın alma ve saklama programları, hedef ürün geliştirme, fiyatlandırmayı özelleştirme ile karlılığı arttırabilirler.
- Kart fiyatlandırma ve karlılık: Kart satıcıları veri madenciliği teknolojisini ürünlerin fiyatlandırmasını karlarını maksimum yapacak ve müşteri kaybını en aza indirecek şekilde kullanarak avantaj sağlayabilirler. Bu yöneme de risk temelli fiyatlandırma olarak adlandırırız.

- Sahtekârlık tespiti: Sahtekârlıkla mücadele çok maliyetlidir. Sahte olduğu belirlenen geçmiş işlemler çözümlenerek, bankalar desenleri tespit edebilir.
- Tahmini yaşam döngüsü yönetimi: Veri madenciliği, bankaların her müşterinin hayat boyu değerini ve her kategoriye uygun şekilde servis verebilmesini (örneğin özel fırsatlar ve indirimler sunmak) tahmin etmesine yardım eder.

3.5.3. Telekomünikasyon

Dünya çapındaki telekomünikasyon şirketleri, artan rekabetle yüz yüze gelmesiyle var olan müşterileri elde tutmak ve yenilerini çekmeyi amaçlayan girişken özel piyasa fiyatlandırma programları üretmeye zorlanmaktadır. Telekomünikasyon içerisindeki bilgi keşfi aşağıdaki gibidir:

- Çağrı kaydı detayı çözümlemesi: Telekomünikasyon şirketleri detaylı arama kayıtları biriktirirler. Benzer kullanım şekillerine sahip müşteri sınıflarını tanımlayarak, firmalar cazip fiyat özellikli özendiriciler geliştirebilirler.
- Müşteri sadakati: Bazı müşteriler rakip şirketlerden cazip teklifleri alarak avantajlardan yararlanmak için defalarca sağlayıcılarını değiştirmektedirler. Firmalar veri madenciliğini kullanarak müşteri karakteristiğini analiz ederek en sadık olanlarını tanımlayabilirler, böylece harcamalarını en çok kar sağlayacak müşteriler üzerine hedeflerler.

3.5.4. Diğer alanlar

Bilgi keşfi uygulamaları daha farklı pek çok alanda da görülmektedir:

- Müşteri sınıflandırması: Tüm sektörler, veri madenciliğinin avantajlarından yararlanarak, geleneksel analizlerin ötesinde ek değişkenleri de dikkate alarak müşteri kaynaklarındaki farklı sınıflandırmaları keşfederler.
- Üretim: Seçim kuralı aracılığıyla, üreticiler müşterileri için ürünleri özelleştirmeye başlamışlardır. Böylece hangi özelliklerin müşteri talebini karşılamak için gerekli olduğunu kestirebilmektedirler.
- Garantiler: Üreticilerin, garanti taleplerini ve bu taleplerin ortalama maliyetini sunacağı müşteri sayısını tahmin etmeleri gerekir.
- Sık uçanların teşvik edilmesi: Havayolları şirketleri daha çok uçmak için teşvik vereceği müşteri gruplarını tanımlayabilirler.
- Güvenlik: Devletler terörist saldırıları keşfetmek için para transferlerini ve iletişim işlemlerindeki benzer örüntüleri tanımlayarak gerçekleşecek yeni olası tehlikeli eylemleri kestirebilmektedirler.
- Trafik güvenliği: Karayollarında kazalara en çok karışan araç türleri ve olayların en çok gerçekleştiği zamanlar tanımlanarak kazalar engellenmeye çalışılmaktadır.

Veri madenciliği uygulamalarına günümüzden gerçek örnekler vermek gerekirse:

Veri madenciliği teknolojisinin en erken uygulamalarından biri olan satış kestirimi yapmak için sinema alanın da Twentieth Century Fox adlı sinema şirketi hangi filmlerin, senaryoların ve aktörlerin hangi eyalet ve bölgelerde daha fazla talep göreceğini kestirmek için gişe makbuzlarını veritabanı olarak kullanarak analiz etmişlerdir (Göral 2007).

Spor dallarında hangi oyuncu/ oyun durumunun yüksek skor için potansiyele sahip olduğunu keşfetmek için Amerika Birleşik Devletleri'ndeki profesyonel basketbol liginde rutin olarak her oyunun detaylı dipnotları çıkartılmaktadır. Bu dip notlarda atış sayısı, pas sayısı kim tarafından hangi zamanda ve hangi oyun düzeninde oynadığı yazılmaktadır. Veri madenciliği uygulaması ile yaratılan ileri keşif sistemi aracılığıyla antrenörlerin fark edemediği bilgiler ortaya çıkmıştır. (örneğin oyuncu x, a bölgesinde oynarken, oyuncu b'nin şut yüzdesinin %75 ten %30 a düştüğünün tespit edilmesi gibi) bu sistem Toronto Raptors ve birçok NBA (Amerikan Ulusal Basketbol Ligi) takımında en iyi oyuncu eşleşmelerini sağlamak için kullanılmaktadır (Hand vd. 2001).

Çizelge 3.2.'de 2006 ve 2009 yılları arasında veri madenciliğinin kullanıldığı alanların sektörel tabanda sonuçları görülmektedir. Bu çizelgedeki araştırmaya 1357 adet şirket katılmıştır.

Çizelge 3.2. Veri madenciliğinin sektörlere göre dağılımları

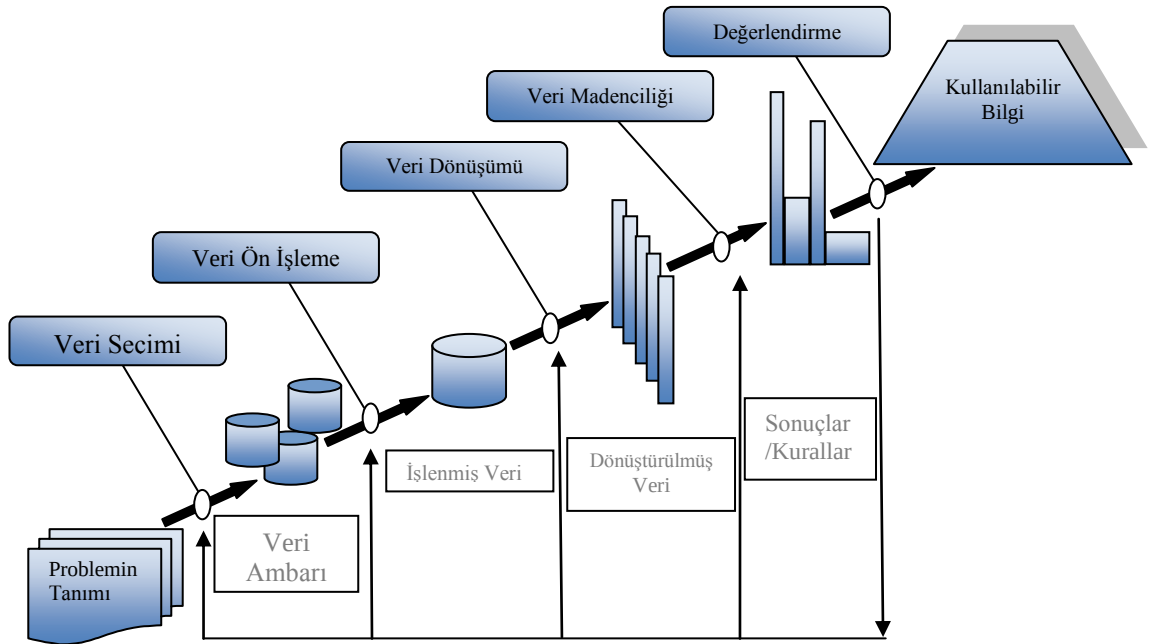
Veri Madenciliğinin Kullanıldığı Sektörler	Dağılımı
Müşteri istekleri yönetimi	%10.17
Bankacılık	%8.40
Perakendecilik	%4.72
Bilim	%4.20
Sahtekarlık tespiti	%5.53
Doğrudan pazarlama	%5.82
Devlet/ Ordu	%2.80
Telekomünikasyon	%5.97
Sağlık/ İnsan Kaynakları	%3.32
Kredi skorlama	%5.01
Diğer	%3.91
Seyahat	%1.55
Eğlence/ Müzik	%1.18
Sosyal politika/Anket araştırmaları	%0.59
Biyoteknoloji/ Genetik	%4.13
Güvenlik/ Anti-terörizm	%1.77
E-posta	%0.44
E-ticaret	%3.46
Web	%5.67
Üretim	%2.28
Sigortacılık	%4.27
Yatırım/ hisse senedi	%2.58

Çizelge 3.2. (devam)¹

Medikal/ eczacılık	%3.24
Finans	%3.76
Çevre ve hayır kurumları	%0.44
Yazılım	%0.44
Yargı	%0.15
Mortgage	%0.22
Eğitim	%0.59
Sosyal ağlar	%1.03
Reklamcılık	%1.40
Kamu hizmetleri	%0.96

3.6. Veri Madenciliği Süreci

Veri madenciliği sürecini bir şekil üzerinde anlatmak gerekirse şekil 3.1. elde edilir.



Şekil 3.1. Veri madenciliği ve bilgi keşfi süreci

3.6.1. Problemin tanımlanması

Bilginin keşfi ve veri madenciliği yöntemlerinin başarılı olabilmesinin en önemli adımı,

¹ <http://www.kdnuggets.com/polls/> 15.02.2010

var olan problemin doğru bir şekilde tanımlanmasıdır. Ortaya konan amaç probleme tam anlamıyla odaklanmalı ve açık bir şekilde tanımlanmış olması gerekir.

3.6.2. Veri seçimi

Veri madenciliğinde kullanılan verilerin boyutu büyük miktarlarda olabilir. Kanser verilerini inceleyen bir bilim adamı 10 yıllık süreçte hastaların özelliklerinden oluşan 10 milyon veriye sahip olabilir. Bu boyutta verinin incelenmesi ve değerlendirilmesi hem uzun zaman hem de büyük uğraş gerektirir. Bu noktada veriler üzerinden örnekleme yapılarak seçim gerçekleştirilir. Hem verinin bütün özelliklerini göstermesi ve veri madenciliği modellerinin uygulanması açısından kolaylık sağladığı için veri seçimi önemli ve öncelikli bir uygulamadır. Ayrıca kurulacak modele göre de veri seçimi yapılabilir. Tahmin edici modelleme yönteminin kullanılması için oluşturulacak eğitim ve deneme verileri, veri seçimi aşamasında yapılır.

3.6.3. Veri ön işleme

Çoğunlukla yapılacak uygulamalar için seçilen verilerin istenilen özelliklere uygun olmadığı görülmüştür. Örnek olarak eksik ve ya aykırı veriler, veritabanında yer almış olabilir. Genellikle veri kümesinde yer alan örneklerin tüm özellikleri değer olarak her zaman tam olmayabilir. Bunun nedenlerini bir anket çalışması üzerinden değerlendirilirse;

- Birey reddettiği için veri alınamamıştır (örneğin birey alkol kullanma durumunu belirtmek istemeyebilir),
- Bireye uygun olmadığı için veri alınamamıştır. (örneğin erkek katılımcı “doğum kontrol hapı kullanıyor musunuz?” sorusunu boş bırakacaktır.)
- Veri alınmıştır ama bilgisayara girilmemiştir. (sekreteryaya hatası).

Veritabanında yer alan tutarsız ve hatalı verilere gürültü; var olan uç verilere de aykırı veriler denir. Örneğin çocuklara ait yapılan bir çalışmada yaş kategorisinde yer alan 30 rakamı aykırı, farklı formatta yazılmış olan 1990 tarihi gürültülü verilerdir. Verilerin ön işlenmesi esasen verilerin temizlenmesidir. Verilerin temizlenmesi kayıp ve gürültülü verilerin ortadan kaldırılması, yanlış ve aşırı uçta bulunan verilerin süzülmesi olarak açıklanabilir. Ayrıca bir örneğe ait özellik iki ayrı sütunda tekrarlı olarak verilmiş olabilir. Bir kişinin yaşı hem sayı olarak hem de doğum tarihi olarak kaydedilmiş olabilir. Bu durumda bir sütunun kaldırılıp özellik sayısının azaltılması gerekir.

Kayıp verilerin yaratacağı sorunları ortadan kaldırmak için aşağıdaki işlemler yapılabilir (Özkan 2008):

- Veri kümesinden eksik verili kayıtların çıkartılması.

Kayıp veriler; veri kümesindeki kayıtların toplam sayısına oranlandığında çalışmanın hassasiyetini etkilemeyecek kadar küçükse bu teknik uygulanabilir. Eğer kayıp veriler büyük bir yekün tutuyorsa bu işlem veri kümesi ile yapılacak uygulamalara zarar verebilir.

- Kayıp verilere tüm verilerin ortalama değerinin girilmesi.

Eksik verilerin yerine sonucu etkilememesi için ortalama değer verilerek çalışmanın yanlış sonuçlara yönelmesi engellenebilir. Örneğin; eksik yaş verilerinin yerine ortalama yaşın girilmesi.

- Kayıp verilere aynı değerin verilmesi.

Kayıp verilerin hepsinin yerine tek değer girilebilir. Fakat bu teknik yapılan uygulamada farklı sonuçlara neden olabilir. Veri madenciliği uygulaması bu boş değeri anlamlı, ayırt edici ve belirleyici bir sonuç olarak çıkarabilir ve algoritmayı, dolayısıyla araştırmacıyı yanıltabilir. Aynı zamanda çıkan bu sonuç araştırma için çok değerli bir bilgi de olabilir. Örneğin etnik kimliğini söylemeyen müşterilerin kayıp değerlerinin yerine “m “ değerinin girilmesi sonucunda m etnik kimliğine sahip müşterilerin belirli bir üründen daha fazla almaları da işletme tarafından araştırılması gereken bir durumdur.

- Kayıp verileri tek tek elle doldurulabilir.

Uygulama için kullanılacak veri kümesi çok küçük olduğunda ve bunun için gerekli zaman olduğunda kayıp veriler el ile tek tek girmek suretiyle doldurulabilir. Bu yöntem genelde kayıp verilerin çalışma için çok önem arz ettiği durumlarda uygulanabilir. Fakat genelde tercih edilmeyen, çok zaman alan bir tekniktir.

- Kayıp verilerin yerine tahmin yöntemi kullanılarak uygun değerlerin verilmesi.

Kayıp verilerin yerine regresyon analizi, beklenti eniyilemesi gibi kestirim yöntemleri ile yeni değerler elde edilip eksik verilerin yerine girilebilir.

3.6.4. Veri dönüşümü

Veri madenciliği uygulamasında kullanılacak teknik ve modeller her veri grubu ile çalışmayabilir. Bazı algoritmalar yalnızca sayısal verilerle çalışırken (örneğin karar destek vektörleri gibi), bazıları da yalnızca sözel (id3 gibi) verilerle çalışabilir. Aynı zaman da her ikisi ile çalışan veya yalnızca 0-1 aralığındaki sayısal verilerle işlem yapan algoritmalar problemin tanımına ve çalışmanın amacına göre kullanılabilir.

Yine çalışılacak verilerin en çok ve en az değerleri arasında çok geniş bir aralık olabilir. Örneğin; karar ağaçları ile yapılan bir çalışmada, çok geniş aralığa sahip veri kümesi aşırı dallanmaya ve yapraklanmaya neden olabilir. Ayrıca büyük ortalamalara ve standart sapmaya sahip değişkenlerin, diğer veriler üzerinde etkisi büyük olabilir. Bu ve buna benzer durumlarda elde edilen verilerin çalışılacak veri madenciliği modeline uygun hale getirilmesi, verilerin farklı bir yapıya dönüştürülmesi zorunludur. Bu dönüşüm sayısal verilerde standartlaştırma ve normalleştirme uygulayarak gerçekleştirilebilir. Bu tekniklerin bazılarını özetlemek gerekirse:

- Min.-maks. normalizasyonu:

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (3.1)$$

Burada X normalleştirilecek veriyi, maks. verinin alabileceği en yüksek, min. verinin alabileceği en düşük değeri göstermektedir.

- Sıfır ortalama normalizasyonu:

$$X^* = \frac{X - \text{ortalama}(X)}{\text{standart sapma}(X)} \quad (3.2)$$

Tüm bu işlemlerin sonucunda veriler algoritmanın uygulanması için hazır hale gelmiştir. Veri önileme genellikle tüm bilgi keşfi sürecinin hem zaman, hem de uğraş olarak %50 ila %80 'ini oluşturmaktadır. Veri hazırlamanın bütünüyle amacı veri madenciliği algoritması için girdi olabilecek veri kümesini oluşturabilmektir.

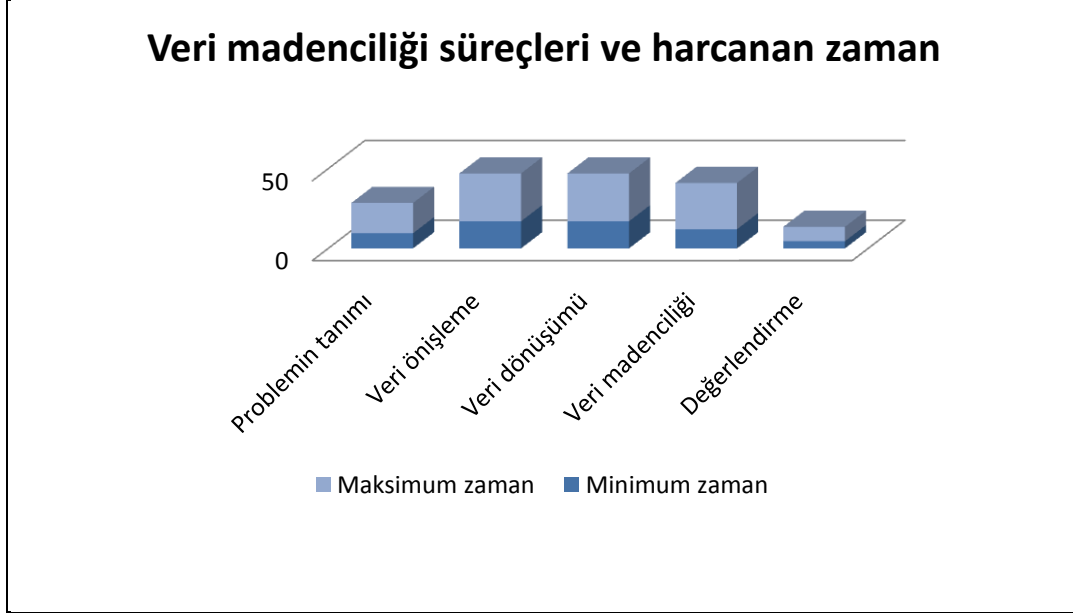
3.6.5. Veri madenciliğinin uygulanması:

Hazır hale gelen verilere veri madenciliği algoritmaları uygulanır. Uygulanacak algoritma, problemin tanımına uygun olarak seçilir ve çözümlenecek veri kümesine uygulanır. Bu aşamada verilerin sorunsuz olması algoritmalarında sorunsuz şekilde çalışmasını sağlayacaktır. Veri madenciliği araçları birçok türde algoritma içerir; bunların arasında bayes metotları, evrimsel algoritmalar, makine öğrenmeleri, sinirsel ağlar v.b. bulunmaktadır (Pal ve Jain 2005). Bu çalışmada kullanılacak olan veri madenciliği araçları ve algoritmalar Bölüm 4'te ayrıntılı şekilde anlatılacaktır.

3.6.6. Sonuçların değerlendirilmesi:

Algoritma uygulanması sonucu, elde edilen sonuçların değerlendirildiği ve anlaşıldığı bölümdür. Bu aşama veri madenciliğinin en önemli aşamalarından biridir. Elde edilen sonuçların doğru bir şekilde yorumlanıp, grafiksel metinlere aktarılması gerekir. Ayrıca elde edilen verinin ne kadar ilginç olduğu ve bilginin etki gücü önemlidir. Eğer ki elde edilen sonuçlar kayda değer bilgiler içermiyorsa veri madenciliği bölümü ve veri önileme bölümü yeniden gözden geçirilerek süreç yeniden denenebilir.

Tüm bilgi keşfi sürecini zaman ve uğraş açısından bir değerlendirmesi yapılırsa şekil 3.2.'yi elde edilir.



Şekil 3.2. Veri madenciliği metotları için gerekli zaman

3.7. Veri Madenciliği Teknikleri

Veri madenciliği, veri tabanlarında bilgi keşfi olarak genel bir tanıma sahip olsa da esasen saklı olan bilgilerin ortaya çıkarılması, daha önce farkına varılmayan fakat potansiyel olarak kullanılması uygun bilgilerin verilerden süzülerek elde edilmesidir. Veri madenciliği büyük veri dizilerine otomatik olarak uygulanan keşif ve gizli özelliklerin ortaya çıkarılması gibi süreçleri kapsar. Genellikle hacmi büyük bu dizilere bu tür uygulamaları elle uygulamak bir hayli zordur. Bu yüzden tüm bu veri madenciliği süreçleri bilgisayarlar ile yürütülmektedir.

İstatistik, yapay zeka ve öğrenme gibi araştırma alanlarının, veri madenciliği ve bilgi keşfi algoritmalarının geliştirilmesine büyük ölçüde katkısı olmaktadır. Bu algoritmaların amacı, diğerlerine nazaran örüntü tanımlama, tahmin, birliktelik oluşturma ve gruplamadır. Bu algoritmalar çeşitli öğrenme problemlerine uygulanabilir. Bazılarının çok iyi sonuç verdiği bir tür problemde, bir başkası çok kötü sonuçlar verebilir ve başarısız olabilir. Bazıları çok iyi yapılanmış veri dizilerinde iyi sonuç verirken, heterojen veri dizilerinde zayıf performans sergilemektedirler. Bazıları yalnızca ayrık verilerde, bazıları sayısal ve sürekli verilerde başarılı olabilmektedir. Belirli özellikler için özel algoritmalar, özellikle belirli alanlara uygun yapılmaktadır; örnek olarak birliktelik kuralı genellikle pazar sepeti analizleri için kullanılır. Naif Bayes belge sınıflandırması için, diğerleri de genel uygulamalar için kullanılmaktadır (Matyja 2007). Bölüm 4'te belirtildiği gibi havacılık kaza verileri heterojen ve ayrık ve sürekli değerler içermektedir.

Veri madenciliği yöntemlerinin çeşitli görevleri vardır:

- Sınıflandırma: Önceden tanımlanan ve ayrık sınıflar, bir dizi örneğe göre daha sonra karşılaşacağımız örnekleri sınıflandırmamızı sağlayan algoritmalar.
- Regresyon (sayısal tahmin): Farklılıklara rağmen sayısal sınıflama yapabilen algoritmalar.
- Birliktelik: Özellikler arasında bağımlılıklarına göre korelasyonlar bulan algoritmalar.
- Kümeleme: Örnekleri, sahip oldukları özelliklere gruplayan algoritmalar.

Sınıflandırma ve regresyon, çoğunlukla tahmin için tercih edilmektedir ve bunun yanında sağladıkları denetimli öğrenme de onların ekstra katkılarıdır. Denetleme özelliği ile hangi algoritmanın öğrenme yaptığı gözlemleyerek anlaşılabilir.

Birliktelik kuralı öğrenmesi sınıflandırma ve regresyonun aksine değişkenler arasındaki ilginç bağımlılıkları bulmaktadır. Birliktelik kuralının, diğerlerinden iki yönüyle farkı vardır. İlki; birliktelik kuralı herhangi bir örneğin herhangi bir özelliğini keşfedebilir ama diğerleri yalnızca bir sınıf üzerinde odaklanır. İkincisi; birden fazla özelliğin tahminini bir kerede yapmak birliktelik kuralı ile olasıdır.

Kümeleme elde edilen örnekleri gruplamak için kullanılır. Genellikle bilgi açısından; bir sınıfın var olmadığı veya birbirinden bağımsız olduğu veri dizilerine uygulanır. Ancak, bu metot sınıflama amaçlı olarak bir özelliğin kayıp değerini belirlemek için bir örneğe yani onu bir kümeye atayarak uygulanabilir. Her küme; birbirine benzer ya da birbiriyle ilişkili olan örneklerin gruplanması ile oluşmaktadır. Kümelerin bölünme kistası, bir probleme dayalı olabilir ve mesafe/ olasılık tabanlı olabilir. Küme gruplarının değerlendirilmesi, genellikle öznel ve kullanıcı için kullanışlı olmasına bağlıdır.

İkisi bu araştırmada kullanılacak olan ve veri madenciliğinin benimsediği üç önemli yöntem bu yöntemlerin metotları şunlardır:

- Sınıflandırma;
 1. Karar ağaçları,
 2. Yapay sinir ağları,
 3. Lojistik regresyon
 4. Naif bayes
- Kümeleme,
- Birliktelik kuralı.

Aşağıdaki alt bölümlerde veri madenciliği yöntemleri ayrıntılı olarak açıklanmaktadır.

3.7.1. Sınıflandırma

Sınıflandırma; verilerin sahip oldukları özelliklere veya farklılıklarına göre kategorilere ayrıştırılmasıdır (Madani 2009). Bir başka deyişle sınıflandırma, sınıfı belli olmayan bir

kayıt için özelliklerine göre tasnif edilmesi sürecidir (Özkan 2008). Bu süreç iki aşamadan oluşur. Birinci aşama; veri sınıflarının veya kavramların karakteristiklerinin tanımlanması için bir model oluşturulmasıdır (bu model oluşturulurken var olan kayıtların öznitelikleri göz önünde bulundurulur). Bu aşamaya, aynı zamanda verilerin eğitimi de denmektedir. Bu eğitimde veriler, veritabanından rastgele seçilmektedir. İkinci aşamada model, gelecekteki verilerin ve nesnelerin sınıflarını tahmin etmek için kullanılır. Açıkçası; ilk grup içinde yapılan araştırma sonuçlarının ikinci gruptaki veriler için de uygulanarak doğrulanmasıdır.

a) Karar ağaçları

Sınıflandırma için birçok teknik kullanılmaktadır. Bu tekniklerinden karar ağaçları; doğruluk, verimlilik ve yorumlanabilirlik açısından en kullanışlı olanı ve en çok tercih edilenidir. Bu metodun avantajlarından biri de kesin olmasıdır. Ayrıca karar ağacı algoritması, diğerlerine göre anlaşılması ve görselleştirilmesi kolay olduğu için en popüler veri madenciliği algoritmalarından biridir (Aftarczuk 2007). Karar ağaçları çeşitli sınıflama görevleri için kullanılır. Sıklıkla kullanıldığı alanların başında; finans, pazarlama, mühendislik, sağlık ve güvenlik vardır (Bölüm 3.5.). Bir karar ağacı, sınıflandırma görevleri için kullanıldığı zaman, sınıflandırma ağacı; regresyon analizi yapmak için kullanıldığı zaman regresyon ağacı olarak adlandırmak uygun olur. Bu çalışmada daha çok sınıflandırma için kullanılan karar ağacından sözü edilecektir. Bu ağaçlar, nesneleri veya örnekleri sınıflandırmak için daha önce tanımlanan sınıfları özellik değerlerini temel alarak kullanır. Mesela sigorta yaptırmak için başvuran bir kişinin “riskli veya risksiz hasta” grubuna sokulabilmesi, bu kişinin yaş ve cinsiyet vb. faktörleri düşünülerek değerlendirilir. Çoğu araştırmacının düşüncesine göre karar ağaçlarının tercih edilme nedeni şeffaflığı ve sadeliğidir. Karar ağaçları, kendiliğinden anlaşılır ve takip edilmesi için çok özel bir veri madenciliği uzmanına gerek yoktur. Karar ağaçları genellikle hiyerarşik yapıdaki grafiklerle gösterilir ki; bu onların diğer tekniklere nazaran daha rahat yorumlanmasını sağlar. Eğer bir karar ağacı karmaşık hale gelirse (çok fazla düğüm oluşursa) o zaman açıkça söyleyebiliriz ki; grafiksel sunum kullanışsız hale gelir. Bir karar ağacı oluşturma süreci yinelenen bir olaydır ve kısa olarak aşağıdaki gibi söz edilebilir.

Karar ağaçları düğümlerden oluşur ve bütün ağacı yöneten düğüm, kök adı verilen kenarları olmayan ana düğümdür. Diğer bütün düğümlerin, yalnızca bir adet giriş kenarı vardır. Bir düğüm, çıkan kenarlarına göre dahili veya test düğümü olarak adlandırılır. Diğer bütün düğümlerde yaprak olarak adlandırılır. (terminal ve ya karar düğümü olarak ta adlandırılabilir). Karar ağacında her yaprak en uygun hedef değeri temsil eden bir sınıfa atanır. Alternatif olarak yaprak, hedef özelliği belirli bir değere sahip olan olasılığı gösteren olasılık vektörü (benzerlik vektörü) tutabilir. Dahili düğümler daireler olarak, yapraklarda üçgen şeklinde gösterilir. İki veya daha fazla dal da her dahili düğümden (yapraktan değil) büyüyebilir. Her düğüm kesin bir karakteristik ile eşleşir ve dallarda değer aralıkları ile eşleştirilir. Bu değer aralıkları verilen karakteristiklerin dizi aralıklarının bölümleridir.

Veriler, ağacın kökünden yaprağına doğru olan yolda yapılan testler sonucu sınıflandırılır. Bir ağacın kökü ile başlanır; bu kök bir niteliğe karşı gelir; verilmiş niteliğin gözlenen değerinin hangi dala uyduğu tanımlanır. Sonra hangi verilen dalın, düğüme karşılık geldiği dikkate alınır. Bir yaprağa ulaşıncaya kadar aynı işlemler, bu düğüm için yinelenir. Karar ağaçları hem sayısal, hem de yazılı nitelikleri birleştirebilir. Her düğüm test edilen nitelik ile ve dallar da onlara karşılık gelen etiketlerle etiketlenir (Rokach ve Maimon 2008). Bu adımlar yeniden tüm özellikler ağaca eklenene kadar yapılır.

Ayrıca karar ağaçları verilerdeki gürültüye karşı çok dirençlidir (Matyja 2007). Karar ağaçlarının yapısal olarak oluşturulması esnasında, önemli bir sorunda verilerin çok fazla eğitilmesidir. Fazla eğitilme genellikle öğrenme evresinin çok uzun tutulduğu veya eğitim örneklerinin az olduğu durumlarda gözlenir. Daha sonra analizcinin, özel rastgele özellikleri eğitim verisine uyarlaması ile tahmin gücünü negatif olarak etkileyebilirler. Bu durum, eğitim örneklerinin performansının arttığı ve görünmeyen örneklerin azaldığı durumlarda gerçekleşir.

Aşırı öğrenme ağacın çok derin olmasına neden olabilir. Bu problemin üstesinden gelmek veya kaçınmak için budama yöntemi geliştirilmiştir. Budama işlemi de fazla ve gereksiz parçaların karar ağacından kaldırılması ile olur (Aftarczuk 2007).

b) Yapay sinir ağları

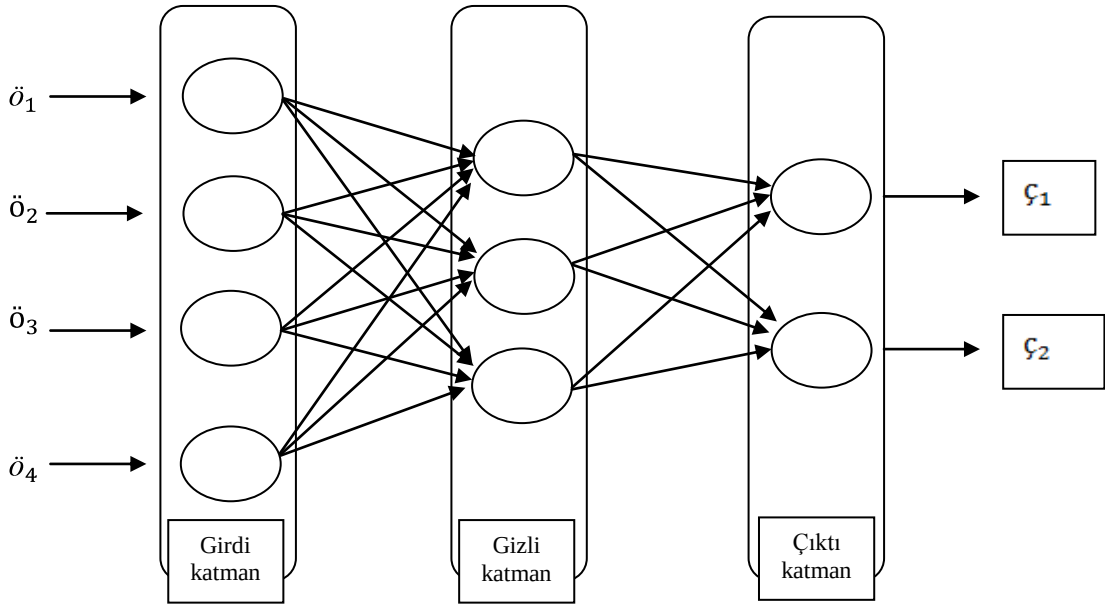
Yapay sinir ağları, insan beyni ilham alınarak üretilmiştir. Yapay sinir ağları insan beyninin gözlenmesine dayalıdır. İnsan beyni nöron ağları ile çok karmaşıktır (Aftarczuk 2007). Nöronlar, basitçe girişleri (birleştirme fonksiyonunu kullanan) birleştirip bir çıktı (aktivasyon fonksiyonunu kullanan) üreten sayısal birimdir. Kısaca yapay sinir ağları basit üç tür nörondan oluşan dizilerle birbirine bağlanmıştır:

- Girdi,
- Gizli ve
- Çıktı nöronları

Ağın yönünü tahmin sürecinde veri akışı simgeler. Özellikler ağın ilk katmanına girdi olarak geçerler. Yapay sinir ağlarına bir örnek şekil 3.3.'te sunulmuştur. Ağlar eğitim verilerindeki hatalara karşı duyarlı değildir fakat öğrenmeleri çok uzun zaman alır. Bu yüzden bu metodun dezavantajı, tahminlerin karmaşıklığı ve zorluğudur. Bununla beraber en büyük avantajları, gelecek verilerin sınıflandırmalarını çok hızlı tahmin etmeleridir. Ancak tahmin mekanizmasının anlaşılması çoğu zaman ciddi sorun teşkil etmektedir (Aftarczuk 2007).

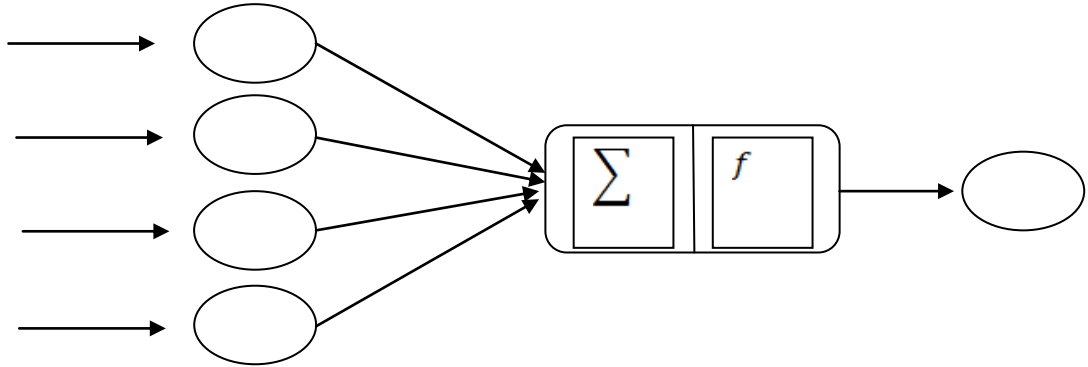
Bir ağda nöronlar tabakalar halinde bulunur. İlki girdi katmanı olarak adlandırılır ve analiz edilecek girdi özelliklerinin değerlerini alır. Bu katmanı gizli nöronların içerisinde olduğu, gizli bir katman takip eder. Bunlar ağa lineer olmayan tahminler yapma imkânı

sunar. Son katman sonuçların elde edildiği çıktı katmanıdır. Verilerin akışı da nöronlar arasındaki bağlantılar boyunca yapılır.



Şekil 3.3. Yapay sinir ağlarına bir örnek (Aftarczuk 2007'den değiştirilerek alınmıştır)

Bu bağlantıların nöronlarda nasıl ilerlediği şekil 3.4.'te sunulmuştur. Her bağlantıya bir ağırlık atanmıştır. Bir ağ düğümler (birimler) ve bağlantılarından oluşmasına rağmen birçok çeşitte ağ bulunmaktadır. Bunlardan ikisi şöyledir; ileri besleme ve ileri besleme olmayan ağlar. Birinci tür ağ, grafik içinde devir olmasına izin vermez, fakat ikincisinde olasıdır.



Şekil 3.4. Bir nöronun gösterimi (Aftarczuk 2007'den değiştirilerek alınmıştır)

Genellikle sinir ağlarının, topolojisini belirlemek kolay değildir. (örneğin gizli nöronların sayısı). Gizli nöronlar daha uzun eğitim zamanı ve karmaşıklık yaratır. Bu sorunun çözümü değişir; uygulamaya ve problemin uygulandığı ağı bağlıdır. Öğrenme sürecinin amacı asıl çıktı ve hedef değer arasındaki kare hata olarak söz edilen çıktı hatalarının tahminini en aza indirmektir. Bu işleme başlamadan önce rastgele tahmin edilen her bağlantıya ağırlıkları atayarak yapılır. Bir sinir ağına öğrenme için birçok algoritma vardır. En popüler olarak kullanılanı geri yayılım algoritmasıdır ve ağıdaki hata aşağıdaki eşitliğe göre hesaplanır:

$$E(\vec{w}) \equiv 1/2 \sum_{\vec{o} \in D} \sum_{k \in \text{çıkıtlar}} (h_{k\vec{o}} - \zeta_{k\vec{o}})^2 \quad (3.3)$$

Çıkıtlar ağıın çıkış nöronlarını gösterir. $h_{k\vec{o}}$ ve $\zeta_{k\vec{o}}$ hedef ve çıkış değerleridir, k 'ncı çıkış nöronu ve örnek D veri dizisinde bulunan eğitim örneği $\vec{o}$ ile ilişkilendirilmiştir (Matyja 2007).

Ağırlıkların atanması için iki adet yaklaşım vardır. Birincisi her örneğin tek başına ağıdan geçtikten sonra değiştirildiği durum (çevrim içi) güncelleştirme yaklaşımıdır. İkincisi de; hata da tüm eğitim kümesi ağıdan geçtikten sonra arka yayılma gösterir. Bu yaklaşıma da devir (yığın) güncellenmesi denir. Öğrenme sürecinin anlaşılması için araştırmacılar hatayı üç boyutlu uzayda bir yüzey olarak hayal etmeyi önerirler. Amaç minimum değeri bulmaktır. Problem ise yüzeyde birden çok minimum olmasıdır. Geri yayılma algoritması olasılıklı, eğimli alçalmalı olarak çalışır ama ne yazık ki küresel minimumu bulmayı garanti edemez.

c) Lojistik regresyon

Lojistik regresyon, iki eşit parçaya bölünmüş problemlere uygulanan istatistiksel bir algoritmadır. Lineer modelleri her bir sınıfa atayan ve oy çokluğuna dayalı modellere dayalı görünmeyen örnekleri tahmin eden iyi bir alternatiftir. Bu yaklaşım dezavantaja sahiptir çünkü bu yaklaşım $[0,1]$ aralığının dışında kalacak çıktı değerlerinin doğru olma olasılığını tam olarak vermez. Lojistik regresyon her gerçek değeri alabilen ve 0 veya 1 ile sınırlanmamış dönüşmüş hedef değişken temelli bir lineer model yaratır. Dönüşüm fonksiyonu lojit dönüşümü olarak söz edilir ve şekil 3.5.'te sunulmuştur. Ortaya çıkan model de yaklaşık olarak aşağıdaki denklem ile bulunabilir:

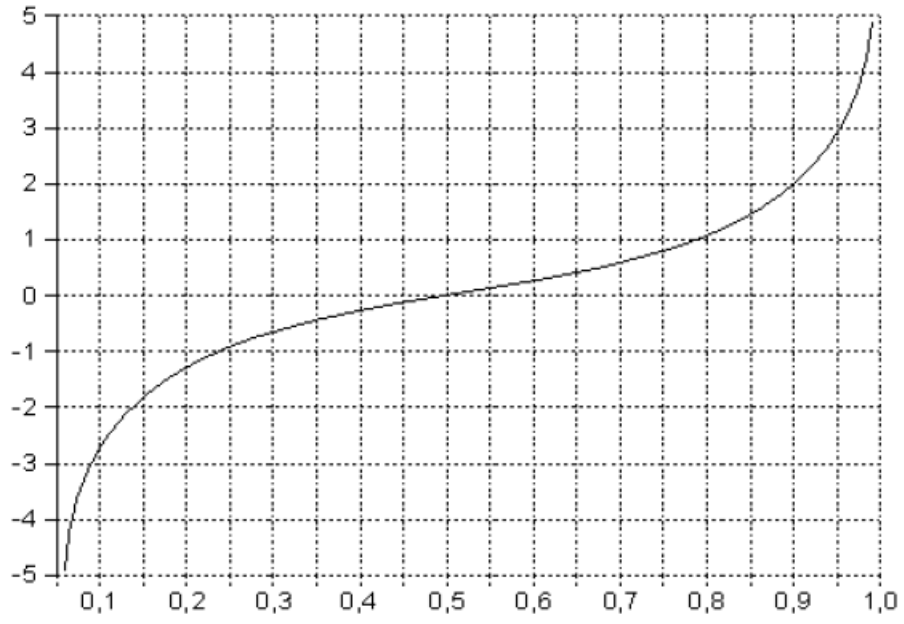
$$\text{Olasılık}[1|a_1, a_2, \dots, a_k] = 1/[1 + \exp(-w_0 - w_1 a_1 - \dots - w_k a_k)] \quad (3.4)$$

w_i belli özelliklerin ağırlığını göstermektedir. Şekil 3.6.'da örnek bir lojistik regresyon fonksiyonunu sunulmaktadır.

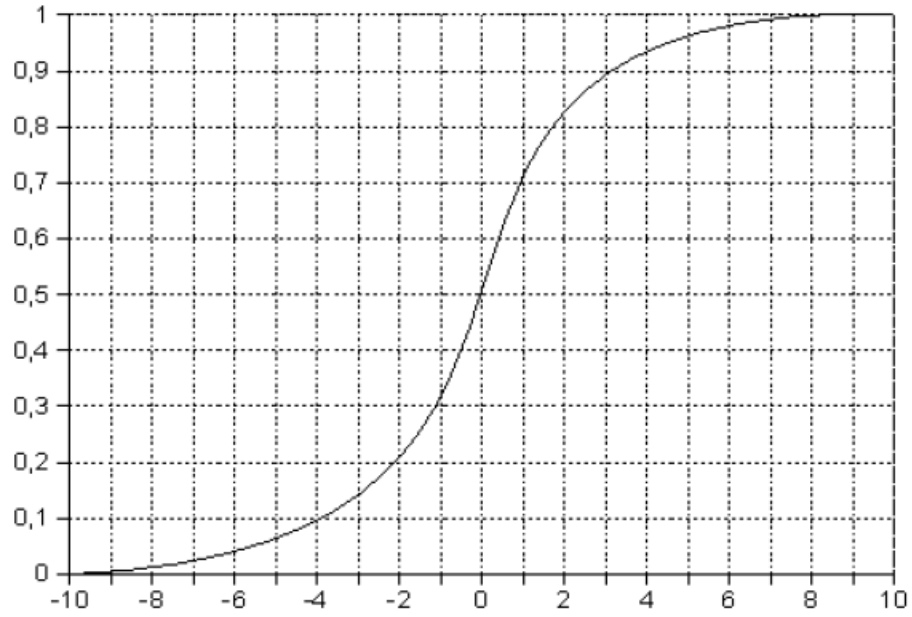
Lojistik regresyon, lineer regresyondaki gibi çok sınıflı durum sorunları ile karşı karşıyadır. Çıkıtlar uygun olasılık değerlerini temsil etmez. Ancak, her çift sınıf değeri için lojistik regresyon modeli oluşturularak bu sorun kolayca çözülebilir. Bu metoda ikili sınıflandırma adı verilir. Her model bir çiftin içinden seçilen gerçek örneklerden oluşturulur. Bu modeller oy çokluğuna dayalı tahminler için kullanılmaktadır.

d) Naif Bayes

Naif Bayes, basit olasılıklı sınıflayıcıdır ve istatistiksel tahmini modellerinin oluşturulması için hızlı bir yöntemdir (Aftarczuk 2007). Algoritması, bilimin diğer alanlarında da kullanılır. Bazen naif bayes algoritmaları diğer gelişmiş algoritmalarından (sinir ağları gibi) daha iyi sonuçlar verir. Araştırmacılar bu yüzden, daha zor ve karmaşık çözümler uygulamadan önce ilk olarak bu algoritmanın denenmesini uygun görmüşlerdir. Algoritmanın saflığı, değişkenleri (özellikleri) bağımsız olarak varsaymasından kaynaklanır. Bu varsayım gerçek yaşam olaylarında doğru değildir,



Şekil 3.5. Lojit dönüşümü (Matyja 2007)



Şekil 3.6. Lojistik regresyon fonksiyonu (Matyja 2007)

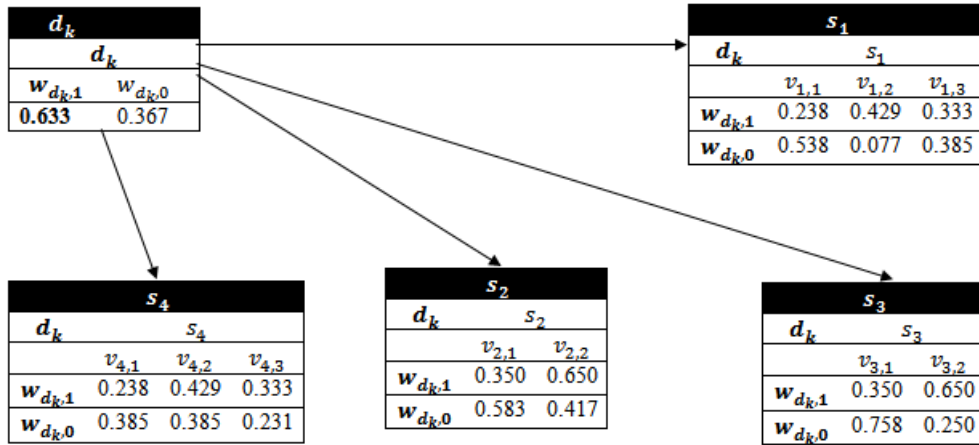
genellikle iyi bir tahmin gücü ortaya çıkarır. Öğrenme süreci; her özelliğin, her değeri ile her sınıfın değerinin sürekli korelasyonlarına dayanmaktadır. Özellik değerleri $(d_1, d_2, \dots, d_n)$ içeren X dizisinin olasılığının T sınıfına ait olması, aşağıdaki eşitlik ile tahmin edilebilir:

$$Olasılık[X/T] \approx N! \cdot \prod_{i=1}^k \frac{o_i^{d_i}}{d_i!} \quad (3.4)$$

Burada $d_1, d_2, \dots, d_n$ ve $O_i^{d_i}$, d_i özelliğinin belirli bir değerinin T sınıf değeri ile korelasyon olasılığını göstermektedir (Matyja 2007).

Ancak bir özelliğin yalnızca bir tek değeri, sınıf değeri ile ilişki içerisinde olmazsa denklem 0' a eşit olur. Bu da demektir ki; sınıfın değeri herhangi bir örnek için her zaman tahmin edilemez. Bu dezavantaj tüm korelasyonlar olası varsayılarak kolaylıkla ortadan kaldırılabılır. Bu çeşitli şekillerde yapılabilir. En bilinen yöntem, bazı katsayıları kullanarak dış bilgiyle modeli birleştirmektir. Varsayım, tüm korelasyonların bir dış bilgi olabilmesidir. Bu yaklaşım sıfır korelasyon değerinde olan örneklerin hepsini bir yaparak modelin sınıf değerinin tahminine olanak sağlar (Aftarczuk 2007).

Naif bayes modelinin yapısı, düğümlerden oluşan bayes ağı formundadır ve her düğüm bir özelliği temsil eder. Düğümler birbirine direkt kenarlar ile bağlıdır ve çevrimsiz bir grafik şeklindedir. Örnek bir bayes ağı şekil 3.7.'de gösterilmiştir:



Şekil 3.7. Bayes ağı (Aftarczuk 2007)

Bayes ağları 1980'lerde geliştirilmiştir. Bayes ağlarının kullanımı olası değişkenler arasındaki etkileşimleri sunmaktadır ve sık sık ilişkilerin neden ve sonucu olarak anlaşılmaktadır.

3.7.2. Kümeleme

Kümelemeyi; kayıtları, gözlemleri ve ya olayları benzer nesnelerin oluşturduğu sınıflara gruplama olarak adlandırabiliriz. Bir küme birbirlerine benzer ve diğer kümelerdekilerle farklı olan kayıtlar topluluğudur. Kümeleme sınıflandırmadan farklıdır çünkü kümeleme için hiçbir hedef değişken yoktur. Kümelemenin görevleri arasında sınıflandırma, tahmin ya da bir hedef değişkenin değerini tahmin etmek yoktur. Bunun yerine, kümeleme algoritmaları göreceli olarak; tüm veri kümelerini benzer kayıtların olduğu kümelerle maksimize edildiği, benzer kayıtların olmadığı kümelerin de minimize edildiği homojen alt gruplara ve kümelere bölmektedir (Larose 2005).

Kümeleme, genellikle veri madenciliği sürecinde bir ön adım olarak yapılır. Sonuçta elde edilen kümeler, ilerde sinirsel ağlar gibi farklı veri madenciliği teknikleri için girdi veri olarak kullanılmaktadır. Muazzam büyüklüğü nedeniyle günümüzde birçok veritabanı; genellikle ilk önce kümeleme analizi uygulanması, daha sonra uygulanacak algoritmaların arama uzayını daraltılmasına yardımcı olmaktadır.

Kümeleme analizi, sınıflandırma analizinde de karşılaşılan birçok benzer sorunu içerir. Bu sorunlara örnek vermek gerekirse:

- Benzerlik nasıl ölçülür?
- Değişkenlerin kategorileri nasıl farklı olarak kodlanır?
- Sayısal değişkenler nasıl standartlaştırılır veya normalleştirilir?
- Kaç tane kümenin ortaya çıkmasını beklenir?

Genel olarak kayıtlar arasında, “Öklid mesafesi” kolay ve rahat uygulandığı için kullanılmaktadır.

$$d_{Euclidean}(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (3.5)$$

Burada “ $x = x_1, x_2, \dots, x_m$ ” ve “ $y = y_1, y_2, \dots, y_m$ ” kayıtların “m” özellik değerini göstermektedir.

Bir başka benzerlik ölçüm yaklaşımı cityblok:

$$d_{cityblok}(x, y) = \sum_i |x_i - y_i| \quad (3.6)$$

Ve yukarıdaki iki ölçütü genel durum ile temsil eden genel bir üs “q” oranı olan “minkowski uzaklığı”:

$$d_{minkowski}(x, y) = |\sum_i |x_i - y_i|^q| \quad (3.7)$$

Kategorik değişkenler için, yeniden farklı fonksiyonlar bir çift kayıttın i’nci özellik değeri için tanımlanabilir.

$$farklı(x_i, y_i) = \begin{cases} 0 & \text{eğer } x_i = y_i \\ 1 & \text{bunun dışında} \end{cases} \quad (3.8)$$

Burada “ x_i ve y_i ” kategorik değerlerdir. Daha sonra $farklı(x_i, y_i)$ Öklid uzaklığının i’nci terimi için kullanılabilir.

En uygun performans için; kümeleme algoritmaları, sınıflandırma algoritmalarına benzer olarak, kullanılacak verilerin normalleştirilmesi gerekir. Böylece belli veriler ve ya değişkenlerin alt kümeleri analizde baskın rol oynamaz. Analizciler ister min-maks normalizasyonu istersede sıfır ortalama normalizasyonu tekniklerini kullanabilir.

Kümeleme algoritmaları hiyerarşik ve hiyerarşik olmayan olmak üzere ikiye ayrılır. Hiyerarşik kümelemede ağaç benzeri küme yapısı (dendrogram); var olan kümelerin tekrarlanan parçalanmaları (bölme) veya birleşmeleri (toplama) ile yaratılır.

Toplama kümeleme metotları, her gözlemi küçük bir küme olarak algılar. İlerleyen basamaklarda, iki en yakın küme birleşerek yeni bir küme oluşturur. Bu yolla veri kümesindeki kümelerin sayısı her adımda birer birer azalmaktadır. Sonunda tüm kayıtlar birleşerek, tek büyük bir küme oluşturur. Bölücü kümeleme metodun da bütün kayıtlar büyük tek bir kümede toplanır, farklı kayıtlar tekrarlı olarak ayrı kümelere ayrılarak her kaydın kendisi bir küme olması sağlanır. Genellikle hiyerarşik kümeleme uygulayan bilgisayar programları toplama metodunu kullanır.

Kümeleme metotlarını uygularken kümeler arasındaki uzaklığı nasıl belirleriz? Ya da iki kümenin birbirine yakın olabilmesi için birbirlerine en yakın komşu mu yoksa en uzak komşu mu olmaları gerekir?

Gelişi güzel A ve B kümeleri arasındaki uzaklığı keşfetmek için çeşitli kriterleri incelemek gerekirse¹:

- Tek bağ, bazen en yakın komşu yaklaşımı olarak da adlandırılır, A kümesinin içindeki en yakın kayıt ile B kümesinin içindeki en yakın kayıt arasındaki minimum uzaklığı temel olarak alır. Bir başka deyişle küme benzerliği, her kümenin elemanlarından birbirine benzer olanlarının bir araya gelmesi temeline dayanır. Tek bağ uzun, ince yapıdaki, bazen de heterojen kayıtların kümelenmesi için kullanılabilir.
- Tam bağ, bazen en yakın uzak yaklaşımı olarak da adlandırılır. A kümesinin içindeki en yakın kayıt ile B kümesinin içindeki en yakın kayıt arasındaki maksimum uzaklığı temel olarak alır. Bir başka deyişle, küme benzerliği her kümenin elemanlarından birbirinden farklı olanların bir araya gelmesi temeline dayanır. Tam bağ, yoğun yapıdaki kayıtların kümelenmesi için uygulanabilir.
- Ortalama bağlantı; çok benzer ve çok farklı kayıtların olduğu gruplarda kümeler arası bağlantı bağımlılığını azaltarak uç değerlerin etkisinden kurtulmak için dizayn edilmiştir. Ortalama bağlantıda, kıstas A kümesi içerisindeki tüm kayıtların ortalama uzaklığı ile B kümesindeki tüm kayıtların ortalama uzaklığı arasındaki uzaklıktır.

Kümeleme algoritmaları arasında en çok kullanılanı K-ortalama kümeleme algoritması; verilerdeki kümeleri bulmak için açık ve efektif bir uygulamadır (Matyja 2007).

¹<http://ninova.itu.edu.tr/en/courses/institute-of-informatics/195/bbl-606/ekkaynaklar?g8396> (11.11.2009)

3.7.3. Birliktelik kuralı

Birliktelik kuralları, bilgi sunumunun bir başka şeklidir ve gerçekleşen durumların birlikte olma olasılığını analiz eden veri madenciliği yöntemidir. Onlar sınıflandırma kurallarını genelleştirerek, yalnızca bir sınıf için değil, her nitelik değeri için tahmin yapmayı sağlar. Bu onlara özelliklerin birleşimlerini de tahmin yeteneği verir. Ayrıca, sınıflandırma kurallarının aksine, birliktelik kuralı dizi halinde değil de tek tek uygulanabilir. Hatta birliktelik kuralı, çok küçük veri dizileri için bile üretilebilir. Birliktelik kuralı çoğu kez müşterilerin satın alma alışkanlıklarının tespiti için kullanılmaktadır. Bu özelliğine de “Pazar sepeti analizi” denir (Akbulut 2006). Bu analiz için kullanılan en yaygın örnek şu şekilde gerçekleşir: Bir müşteri, A nesnesinden sürekli ama aynı zamanda B ürününü de ara sıra alıyor. Fakat aynı zamanda bu müşterinin, A ve B ürününü aldığı zaman aynı zamanda C ürünü alma olasılığının hesaplanmasıdır. Birliktelik kuralları, sayısı çok büyük olmasına rağmen, birçok bilgi ile alakalı olmayabilir. Hangi kuralın potansiyel olarak en çok bilgiyi içerdiğini belirlemek için iki ölçüt tanımlanmıştır. Bunlar:

- Kapsam: Destek sayısı olarak da adlandırılır, kuralın uygulandığı veri sayısını gösterir. Bir başka deyişle A,B nesnelerinin satın alınmasının tüm satın alma işlemleri içersinde kaç kere yinelandığını gösterir.
- Doğruluk: Güven ölçütü olarak da adlandırılır. Kuralın uygulandığı veri sayısına oranla kaç tane özelliğin doğru olarak tahmin edildiğinin söz edilmesidir. Bir başka deyişle A nesnesini alan müşteri grubunun, B nesnesini de aynı zamanda alma olasılığını gösterir (Özkan 2008).

Birliktelik kuralı ve karar ağaçları onları yaratan algoritmanın türü açısından benzerdir. Hem ağaçlar hem de kurallar, eğitim dizisini parçalara ayırarak yapılan böl ve yönet türünde algoritmalar kullanılmaktadır. Ancak sunum açısında farklar vardır. Birliktelik kuralları, insan bakış açısından anlaşılması çok kolaydır. Matematiksel dilde, birliktelik kuralını açıklamak gerekirse:

$N = \{N_1, N_2, \dots, N_n\}$ dizisi satın alınan nesnelerin kümesi olsun.

$I = \{I_1, I_2, \dots, I_m\}$ dizisi veritabanında alışveriş sonucu kaydedilen işlemleri gösterebilir.

Her bir işlemin alacağı sayısal değer 1 veya 0'dır. Bir başka deyişle alışveriş olmuş ya da olmamıştır. Buradan eğer $I_1 = 1$ koşulu varsa N_1 satın alınmıştır demektir. Y kümesi de N kümesinin bir alt kümesi olsun. Bu tanımlamalardan sonra birliktelik kuralı şu şekilde izah edilebilir:

$Y \Rightarrow N_k$, N_k N dizisi içerisindeki herhangi bir işlemdir fakat bu eleman Y alt kümesi içerisinde değildir. $Y \Rightarrow N_k$ kuralının, I dizisindeki bir işlem olup olmadığını bulmak için güven aralığının olması gerekir. Bu da % g ile gösterilir ve ilişkisel olarak da gösterimi:

$$Y \Rightarrow \frac{N_k}{g}, 0 \leq g \leq 1 \text{ 'dir.} \quad (3.9)$$

Birliktelik kuralının uygulanması için birçok yöntem kullanılmaktadır. Bunlardan en yaygın kullanılanı “apriori algoritmasıdır”. Bundan farklı olarak da “Fp-growth, Predictive apriori, AIS, AprioriTid ve SETM” algoritmaları da bulunmaktadır (Silahtaroglu 2008).

3.8. Veri Madenciliği Yaklaşımının Çözümleme İçin Benimsenmesinin Nedenleri

Çalışmanın bu bölümünde veri madenciliği yöntemi ayrıntılarıyla anlatıldıktan sonra verilerin çözülmesi için neden tercih edildiği açıklanmıştır. Bu konu irdelenirken uçuş verilerinin karmaşık yapısı ve her uçuş verisinin kendi başına özel olması göz önünde bulundurulmuştur. Çalışma için veri madenciliği yönteminin tercih edilme nedenleri aşağıdaki gibidir;

3.8.1. Var olan uçak kaza verilerindeki kısıtlamalar

Havacılık güvenliği verileri, esas olarak kaza ve uçak olaylarını içeren bilgileri içerir. Güvenlik verileri sorunsuz uçuşlardan oluşmaz. Yani, bu çalışmada kullanılan veriler yalnızca sorunlu uçuşlara ait bilgilerdir. Bu nedenle bu çalışmada kullanılan veriler, güvenli ve güvensiz uçuşların karşılaştırılması değildir. FAA veritabanı Amerika Birleşik Devletleri’nde gerçekleşen tüm sivil uçak kaza ve olayları ile ilgili raporları içerir. Kazalar, bağımsız bir kuruluş (FAA kurulu) tarafından incelenmektedir. Bu yüzden kaza verileri ön yargıdan uzaktır. Aynı kanıyı uçak olayları için de söyleyebiliriz; çünkü FAA/AID veritabanı uçak olaylarına çok daha fazla önem verir ve inceler.

Kullanılan veritabanındaki veriler sayısal ve sözel olarak veya hem sayısal hem de sözel nitelikler göstermektedir. Ayrıca veriler arasında eksik veri girişleri ya da aykırı/gürültülü girişler mevcuttur. Bu genellikle tam anlamı ile intizam içinde tutulmamış veritabanlarında sık sık gerçekleşen ve beklenen bir durumdur. Bunun nedenleri arasında, verilerin bazen ulaşılamaz olması veya veri toplanan örneklerin hepsinin tüm özellikleri bünyesinde bulundurmamasıdır. Örneğin bir ülkenin milli gelirinin nüfus üzerine dağılımını incelerken oluşturulan veri tabanında, kişilerin maaşı bölümüne bir ülkenin bireyi olan çocuklara karşılık gelen kısım mecburen boş bırakılmak zorundadır. Aynı şekilde çalışmada da bazı uçak modellerinin fazla tanınmamış olması ve kullanılan formlarda bulunan özelliklerin hepsine sahip olmamasından dolayı eksik veriler bulunmaktadır. Bu da verilerin kullanılmasında kısıtlar oluşturmaktadır.

3.8.2. Çözümleme için diğer istatistiksel tekniklerin uygunluğu

Kesin istatistiksel teknikler (çoklu regresyon ve doğal korelasyon analizi gibi) sivil uçak kazalarının çeşitli özellikleri arasındaki ara bağlantıları ortaya çıkarmak için değerlendirilmiştir. Çoklu regresyon, bir kriter değişkeni (bağımlı) için iki veya daha fazla tahmin yapabilir. Değişkenlerin (bağımsız) bilinen değerlerine dayalı bir tahmin değeri de üretebilir. Doğal korelasyon ise; çoklu regresyonun genişlemiş hali olarak kabul edilen halidir; iki takım çok değişkenli değişkenleri aynı anda uğraşma yeteneği vardır ve hem yapısal hem de uzaysal anlamlar üretebilir.

Değerlendirme üzerine, istatistiksel tekniklerin uçak kaza verilerini çözümlemek için uygun olmadığı sonucuna varılmıştır. Bunun nedenleri arasında önemli olan iki tanesi şunlardır; ilk olarak istatistiki yöntemler için aralık düzeyinde veri gereklidir. Eğer sıra düzeyinde veri analiz edilmesi gerekiyorsa, her sıra düzeyinde özellik için sözde değişkenler yaratılmalıdır. Böylece sıra düzeyinde, veri aralık düzeyinde veri olarak görülür. Uçak kazaları veritabanı da sıra düzeyinde büyük sayıda veri içerdiği için bu veriler içinde büyük sayıda sözde değişkenler yaratılarak aralık düzeyinde veriye dönüştürülmelidir. Fakat bu işlemi yönetmek son derece zordur ve anlaşılabilir sonuçların elde edilmesi oldukça zaman alan bir işlemdir. İkinci kriter ve tahmin değişkenleri makul tanımlı matematiksel fonksiyonlarla ilişkili olmalıdır. Ayrıca uçak kazası veri özelliklerinden hangilerinin bir diğeri ile ilişkili olduğu çözümlemeye önce bilinmemektedir.

İstatistiksel analiz ve veri madenciliğinin karşılaştırması/ farklılaştığı noktalar çizelge 3.3.'te genel hatları ile yer almaktadır.

Bu noktada vurgulamak gerekir ki, Bölüm 3.8.1.'de belirtilen nedenlerden ötürü ve verinin aşırı büyük olması, yüksek boyutu, homojen olmaması, dağınık yapısından dolayı istatistiksel çözümlemeler bu çalışma için yararlı amaçlara hizmet edemez.¹

İstatistiksel çözümlemelerin, araştırma problemleri için kullanılması o problemin türüne bağlı olarak değişmektedir. Örneğin, genel bir kural olarak, uçak kaza verilerinin kapsamlı eğimi araştırılacaksa, geleneksel istatistiksel yöntemler kullanılabilir. Eğer bununla birlikte çeşitli özellikler arasındaki ara bağlantılar araştırılacaksa, veri madenciliği teknikleri kullanılmalıdır.

Buna ek olarak, veri madenciliği disiplinler arası bir uygulamadır. İstatistik, veritabanı teknolojileri, makine öğrenmesi, model tanımlama, yapay zeka ve görüntüleme gibi metodlar veri madenciliği uygulamasında rol oynar. Bu yüzden bu disiplinler ile veri madenciliği arasında keskin bir sınır tanımlamak çok zordur. Ayrıca bir kişinin veri madenciliği problemi; bir başka kişi için istatistik, veritabanı veya makine öğrenmesi problemi olabilir (Hand vd. 2001).

Çizelge 3.3. İstatistiksel çözümleme ve veri madenciliğinin karşılaştırılması
(Koyuncugil ve Özgülbaş 2009'dan değiştirilerek alınmıştır)

İstatistiksel Çözümleme	Veri Madenciliği
İstatistikçiler genellikle bir hipotez ile başlarlar.	Veri Madenciliği hipoteze gerek duymaz
Hipotezlerini eşleştirmek için kendi eşitliklerini geliştirmek zorundadırlar.	Veri Madenciliği algoritmaları eşitlikleri otomatik olarak geliştirir
İstatistiksel çözümlemeler yalnızca sayısal verileri kullanır.	Veri Madenciliği farklı tiplerde veri kullanır (örneğin metin, tarih, metin ve sayı). Yani yalnızca sayısal veriyi değil.
Kirli veriyi analizleri sırasında bulur ve süzer.	Veri Madenciliği temiz veriye dayanır.
İstatistiksel Analiz sonuçlarını yorumlamak daha kolaydır.	Veri Madenciliğini sonuçlarını yorumlamak kolay değildir. Sonuçlarını analiz etmede ve yorumlamada ve bulguları yöneticilere iletmede mutlaka istatistikçiye gereksinim duyulmaktadır.

Bir sonraki bölümde uçak kaza verileri veri madenciliği teknikleri ile detaylı olarak çözümlenecek ve tartışılacaktır. Sonuçların ışığında; kazaların oluşmasına neden olan çeşitli etmenler, özellikle de insan hataları ile tanımlanacaktır. Böylece hava yolu güvenliği için çeşitli formüller elde edilerek tavsiyelerde bulunulacaktır.

¹http://fazliyildirim.com/pdf/364_3.pdf (08.10.2009)

4. UYGULAMA

4.1. Giriş

Mitra ve Acharya (2003) veri madenciliğini; veri tabanlarında bilgi keşfi ile eşanlamlı olarak ve de geçerli, yeni, potansiyel olarak yararlı ve en sonunda anlaşılabilir desenleri verilerden elde edebileceğimiz bir süreç olarak tanımlamaktadırlar. Ayrıca toplam sürecin alçak-düzyer veriyi yüksek düzeyde bilgiye dönüştürmeyi sağladığını belirtmişlerdir.

Veri madenciliği; hissedilebilir bilgiyi çıkarmak ve yüksek hacimde veriden ilginç ilişkileri oluşturma, egemen olguları açıklamak, karar verme sürecini desteklemek ve gelecekteki olayları tahmin etmek için güçlü araçlar sunmaktadır. İsminden de anlaşılacağı gibi, veri madenciliği teknikleri, veri tabanlarından bilgiyi kazıp çıkarmaktadır. Analiz için kullanılan veritabanı aşağıda tanımlanmıştır.

4.2. Analizde Kullanılan Veritabanı

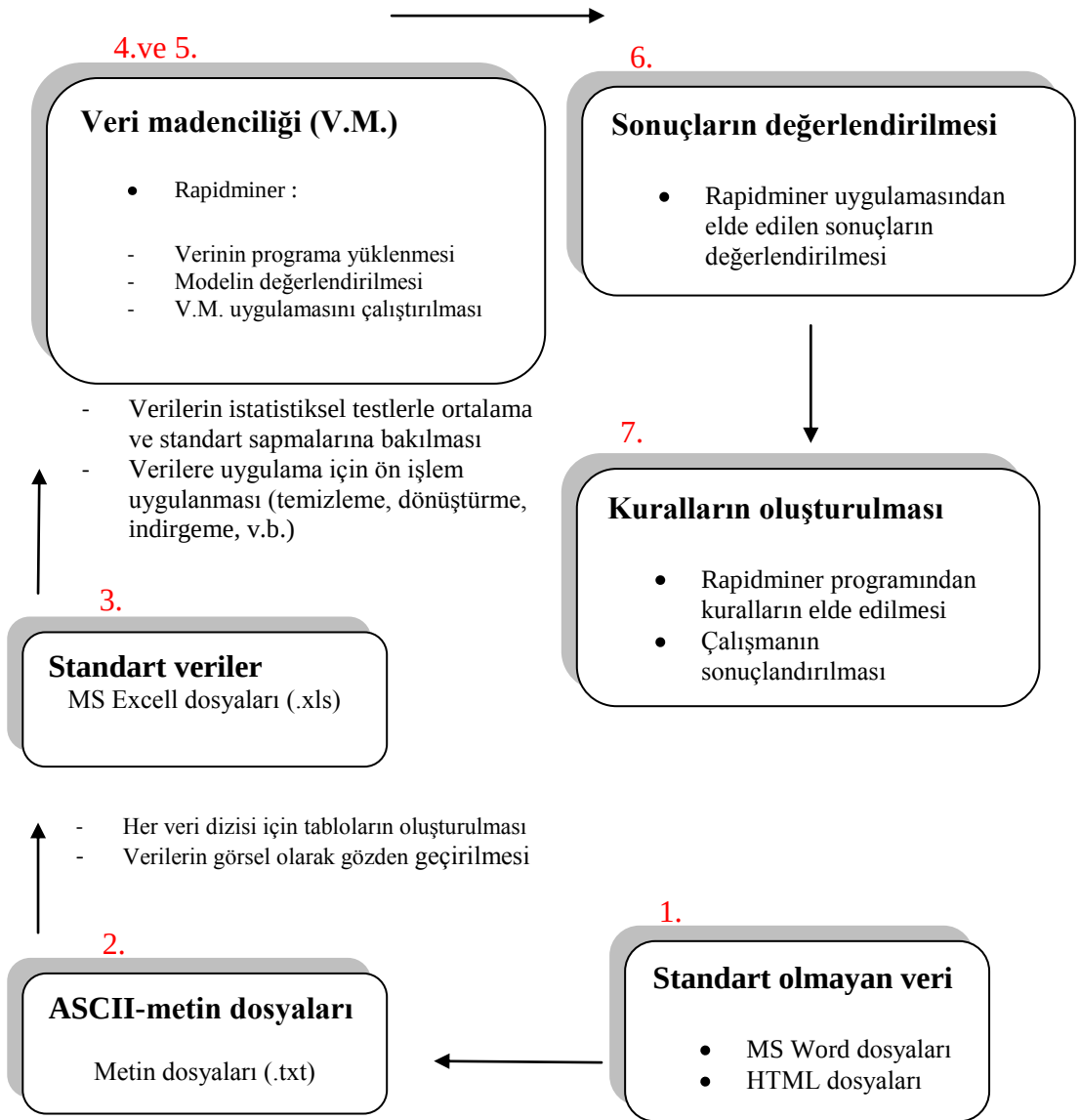
Çalışmada kullanılan veriler, Amerikan Federal Havacılık Kurumu kaza/ olay veri sisteminden alınmıştır. FAA havacılık güvenliğinin sürekli olarak artırılması için kaza/olay verilerini açık kaynak olarak sunmaktadır. Bu temel amaç için FAA Havacılık Güvenliği Analiz ve Paylaşım Sistemi'ni (ASIAS) kurmuştur. ASIAS, kullanıcıların çeşitli veritabanları üzerinde birleşik sorgulamalar yapmasına, geniş veri ambarlarında arama yapmasına ve uygun elemanların kullanışlı formatlarla kullanılmasına olanak sağlar. Kaza/ olay veri sistemi, 1978 yılından bu yana genel havacılık ve ticari hava taşımacılığı kaza ve olaylarının raporlarını içermektedir. Veritabanı, kazalara ait raporları Amerika Birleşik Devletleri Ulusal Ulaştırma Güvenliği Kurumu'ndan (NTSB) direkt olarak alır. Onun dışında uçak olayları direkt olarak Form 8020-5'ten alınan veriler ile oluşturulmaktadır (Gülbüz vd. 2009). Uçak olaylarını kısaca uçağın herhangi bir hasar almadığı ve yolcuların yaralanmadığı uçak kazaları olarak adlandırabiliriz. Bu çalışma için, 2005 - 2010 yılları arasındaki veriler bu sistemden elde edilmiştir.

Çizelge 4.1.'de alınan verilerin çalışma esnasında izlediği yol anlatılmış. Verilerin ilk hali metin belgesi (.txt) ve MS Word. rapor formatında AID sisteminden alınmıştır. Veri tabanında bulunan 164 adet özellik veri ön işleme yöntemi ile ayıklanacaktır.

4.3. Yazılım Uygulamaları

Bu çalışmada Rapidminer 5.0.6 programı; 2005'ten 2010'a kadar uçak kaza verilerinin çözümlemesi için seçilmiştir. Rapidminer programı¹, Amerika Birleşik Devletleri'nde bulunan YALE Üniversitesi bilim adamları tarafından Java programlama dili kullanılarak geliştirilen bir veri madenciliği programıdır. Bu program, birçok dosya türünü destekleyerek, veriler üzerinden anlamlı bilgiler çıkartabilir. Üzerinde çalışabildiği dosya uzantılarına örnek vermek gerekirse; aml, arff, att, bib, clm, clx, cri, csv, dat, ioc, log, xls bunlardan bazılarıdır. Diğer programlara nazaran birden çok dosya türünü desteklemesi Rapidminer programının artılarından. Rapidminer programının;

Çizelge 4.1. Verilerin çalışma esnasında izlediği yol



¹ <http://rapid-i.com/content/view/167/82/> (10.10.2009)

sınıflandırma ve regresyon, karar ağaçları, naif bayes, mantıksal kümeler, ilişkilendirme kuralları ve kümeleme için birçok algoritma (k-means, k-medoids,), veri önileme için ayırma, normalizasyon, filtreleme gibi özellikler, genetik algoritma, yapay sinir ağları, 3D ile verileri analiz etme gibi birçok özelliği bulunmaktadır. Bünyesinde 400’den fazla algoritma barındırmaktadır. Bu program, kuruluşlara ve araştırmacılara ellerindeki verilerin derinlemesine çözümlenmesini ve arka planında ne tür ilişkilere sahip olduğunun anlaşılmasına yardım eder. Ayrıca bir veri madenciliği aracı olarak gelişmiş modelleme teknolojisi ile kolay kullanım olanağı sağlar, kuruluşların veri içerisindeki ilginç ve değerli ilişkileri keşif ve tahmin etmesini sağlar. Kuruluşlar ve araştırmacılar bu nedenle Rapidminer 5.0.6 karar verme sürecindeki kaliteyi arttırmak ve değerli bilgileri daha iyi ve derin bir perspektif içinde yönetmeyi sağlar.¹

Çalışmada Rapidminer 5.0.6 programının kullanılmasındaki başıca nedenleri sıralamak gerekirse:

- Program 20’ye yakın dosya biçimi ile veri girişi yapılmasına izin vermektedir.
- Program diğer veri madenciliği programlarında kullanılan algoritmaları da içermektedir. (Weka programında kullanılan veri madenciliği algoritmaları gibi)
- Program açık kaynak kodlu olarak yayınlanmakta ve içerisindeki modüller, sık sık bir üst sürüm ile güncellenmektedir. Program kurucuları kurdukları forum sitesinden, kullanıcılardan gelen programın hata mesajlarını ve sorunlarını sürekli değerlendirerek, yardım ve geliştirme hizmeti sunmaktadırlar.
- Programın veri önileme bölümü gelişmiş ve kullanıcının kullanım yeteneklerine göre şekillenebilmektedir.
- Programdan alınan sonuçların aktarımı birçok şekilde ve kolaylıkla olmaktadır.
- Program üzerinde hali hazırda bulunan algoritmalar, uzman kullanım seçeneği ile farklılaştırılabilmektedir.

4.4. Uygulamada Kullanılan Algoritmalar²

Rapidminer programı tarafından sınıflandırma ve birliktelik kuralları altında birçok algoritma hazır halde bulunmaktadır. Fakat bu algoritmalar programın içersinde değişik isim ve şekillerde sunulmaktadır. Bu yüzden çalışmamızda kullanılan algoritmaların literatürde geçen algoritmalarla olan ilişkisi, gerçek adları ve programda belirtilen isimleri ile çalışma prensipleri bu aşağıda anlatılmıştır;

¹ <http://garr.dl.sourceforge.net/project/yale/1.%20RapidMiner/4.6/rapidminer-4.6-tutorial.pdf> (30.05.2010)

² (Özkan 2008, Silataroğlu 2008)

4.4.1. Karar ağacı algoritmaları

Sınıflandırma yönteminin en çok kullanılan tekniklerinden karar ağaçları birbirlerinden, ağacın nasıl oluşturulduğuyla ayrılır. Bu bağlamda çalışmada kullanılan algoritmalar aşağıdaki gibidir:

a) İD3 algoritması

J. Ross Quinlan tarafından geliştirilen algoritmalarından biridir. Bu algoritmanın diğer sınıflandırma algoritmalarından farkı, dallanma kriteri olarak entropi kavramını kullanmasıdır. Yani sınıflandırma yapılırken ayırt edici özelliği tespit ederken entropiden yararlanır. Entropi kavramından söz etmek gerekirse; herhangi bir sistemde var olan belirsizliğin boyutuna entropi denir. Entropi rastgeleliği, belirsizliği ve beklenmeyen durumun ortaya çıkma olasılığını gösterir. Eğer bir sistemde bütün veriler tek bir sınıfa ait olursa entropi 0 olur. Entropi kavramı 0 ile 1 arasında değişkenlik gösterir. Entropinin maksimum olabilmesi için sınıflara ait olasılıkların hepsinin eşit olması gerekir. Entropiyi matematiksel olarak söz etmek gerekirse;

Bir T veri dizisine ait olasılıklar $P_T = \{p_1, p_2, \dots, p_n\}$ olsun. Bu veri dizinin entropisi

$$E(T) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (4.1)$$

denklemleri ile gösterilir.

- Karar ağaçlarının entropiyi kullanması:

Karar ağaçlarının oluşması esnasında, dallanmanın hangi özellikten başlayacağını karar ağacı algoritması belirler. Dallanmanın başlama noktasının seçilmesi zordur ve önemlidir. Örneğin beş özellik ve 20 elemanlı küçük bir veri kümesinde bile, elde edilebilecek karar ağacı sayısı 10^6 'dan büyüktür (Özkan 2008). Bu yüzden dallanmanın seçilmesi için, algoritmalar türüne göre hesaplama yapar. İd3 algoritması da belirsizliğin alacağı değere göre dallanmaya yön verir. İd3 algoritması bütün özelliklerin belirsizliğini hesaplar ve hedef özelliğin belirsizlik değerinden çıkarır. Elde edilen bu değer kazanç değeridir. En yüksek kazancın elde edileceği özellikten dallanmayı başlatır.

b) C.4.5 algoritması

Çalışma için kullanılan Rapidminer programında geçen adı W-J48'dir. J.Ross Quinlan tarafından 1993 yılında, Id3 algoritmasının eksiklerini tamamlamak amacıyla geliştirilmiştir (Quinlan 1993).

- C 4.5 algoritması farklı olarak eksik verileri, karar ağacı oluştururken kullanmaz.
- Id3 algoritmasında gerçekleşen aşırı dallanma durumunu engellemek için, karar ağaçlarında ön budama ve budama tekniklerini kullanır. Budama işlemi için iki teknik kullanır. İlki; alt ağaç yerleştirmesi için ağaç içindeki alt ağaçlar yapraklara dönüştürülür. Bu değişimin yapılabilmesi için konulacak yaprağın

hata değerinin, alt ağacından düşük olması gerekir. İkincisi ise; hata değeri değerlendirildikten sonra bir alt ağacın, bu ağacı en fazla kullanan ağacın yerini almasıdır.

- C 4.5 algoritması, Id3 algoritmasında gerçekleşen aşırı öğrenmeyi önlemek için yeni bir kazanım oranı üretmiştir.

Kazanım oranı = Kazanım/ Ayırma bilgisi (entropi)

c) Desicion Tree Algoritması

Bu algoritma Quinlan'nın C 4.5 algoritmasına benzer şekilde çalışır. C 4.5 algoritmasından farklı olarak dallanma kriteri olarak entropi ve kazanım oranını değil, bilgi kazanımını kullanmaktadır.

Bilgi kazanımında matematiksel olarak söz etmek gerekirse;

T ve L veri dizileri olsun. Bu veri dizilerinin bilgi kazancı:

$$B(T,L)= E(L)+E(T)-E(L,T) \text{ dir.}^1 \quad (4.2)$$

Desicion Tree algoritması C 4.5 algoritmasından farklı olarak üç kriterle bağlı olarak durmaktadır²:

- Dallanma esnasında veri dizisinde bulunan özelliklerden hiçbiri, eşik bilgi kazancı değerini geçemezse,
- Maksimum karar ağacı genişliği aşılsa,
- Veri dizisindeki örnek sayısı minimum dallanma sayısından az ise.

d) CHAID Algoritması

Desicion tree algoritması ile benzer şekilde çalışır fakat farklı dallanma kriteri olarak bilgi kazancı veya entropiyi değil ki kare dağılımını kullanır.

4.4.2. Birliktelik kuralı algoritmaları

a) FP-growth algoritması

FP-growth algoritması kısaca, sık yinelenen örüntülerin birbirlerini bularak birliktelik kurallarının oluşturulmasını sağlar. Bu algoritmanın diğer birliktelik algoritmalarından farkı, tüm veri tabanını küçülterek yoğun bir veri yapısı olan, sık örüntü ağacına dönüştürmesidir. Genel olarak kullanılan Apriori algoritmasından farklı olarak, tüm veri

¹ http://ftp.utcluj.ro/pub/users/nedeveschi/AV/5_EvaluarClasificatori/entropyEtc.pdf (03.03.2010)

² <http://garr.dl.sourceforge.net/project/yale/1.%20RapidMiner/4.6/rapidminer-4.6-tutorial.pdf> (30.05.2010)

tabanını, oluşturulan n aday sayısı kadar değil de yalnızca iki kez tarar ve sonuca ulaşır. Bu sayede çalışma hızı yüksektir, az kaynak tüketir ve büyük veri tabanları için kullanımı idealdir. Algoritma aşağıdaki adımları takip ederek sonuca ulaşır:

- Veri tabanı taranarak her veri için, bir destek değeri oluşturulur.
- Veriler içerisinden, destek değeri kullanıcının belirlediğinden küçük olanlar veri tabanından çıkarılır, diğerleri destek değerine göre büyükten küçüğe sıralanır.
- İkinci taramadan önce yinelenen örüntü ağacının oluşması için bir düğüm oluşturularak, kök elde edilir ve tarama başlatılır.
- Taranan veriler sırayla ağaca yerleştirilir. İlk defa yerleştirilenler bir düğüm oluşturur ve destek değeri bir olarak belirlenir. Tarama esnasında karşılaşılan eş veriler için yeni bir düğüm oluşturulmaz, var olan düğümün destek değerine bir eklenir.
- Büyük veri tabanları üzerine uygulandığından, ağacın dalları arasında daha kolay dolaşmak için bir başlık çizelgesi oluşturulur. Her verinin ilk yerleştiği düğüm, bu çizelge üzerine işaretlenerek aradaki bağların durumu belirlenir.
- Oluşturulan sık örüntü ağacı üzerinde, en düşük destek değerine sahip düğüm üzerinden başlayarak, algoritma her düğüm için çalıştırılır. Düğümlerin destek değerlerine göre, elde edilen yollar kural olarak oluşturulur.

b) Apriori algoritması

Çalışmada kullanılan Rapidminer programında, W-apriori olarak adlandırılır. Agrowal ve Srikant tarafından 1994 yılında geliştirilen bir algoritmadır. Geniş veri kümelerinden bağlantı analizinin yapılması ve bu bağlantı kurallarının ortaya çıkarılmasında kullanılan popüler bir algoritmadır (Özkan 2008). Algoritma aşağıdaki adımları takip ederek sonuca ulaşır:

- Birliktelik bağlantılarının ortaya çıkarılmasında destek ve güven ölçütlerinin kıyaslanması için en başta kullanıcı tarafından eşik değeri belirlenir. Algoritma çalışırken elde ettiği sonuçların mutlaka eşik değerinden büyük veya eşit olması gerekir.
- Algoritma veri tabanını tarayarak her biri veri için bir satır oluşturur ve destek ölçütleri hesaplar. Bunu da her bir verinin kaç kez yinelandığını sayarak yapar. Her veri satırının destek değeri eşik değeri ile karşılaştırılır. Verilerden destek değeri eşik değerinden büyük ve eşit olanlar dikkate alınır, diğerleri bağlantı kurallarından çıkarılır.
- İkinci adımda elde edilen veri satırları bu adımda ikiyeşerli olarak gruplandırılır ve destek değerleri hesaplanarak eşik değeri ile karşılaştırılır. Kurala uyan satırlar alınır diğerleri silinir. Bu uygulama aynı şekilde üçlü, dördü, v.b. içinde çalıştırılır.
- En sonunda artık gruplanacak veri kalmadığı zaman kural destek ölçütleri eşik değeri ile test edilir ve kurallar üretilir. Daha sonra da her bir kural için güven ölçütü elde edilerek algoritma tamamlanır.

c) Predictive (Tahminleyici) Apriori algoritması

Çalışma da kullanılan, Rapidminer programında geçen adı W-predictive apriori'dir. Bu algoritma, Tobias Scheffer tarafından geliştirilmiştir. Predictive apriori algoritması, birliktelik kurallarını elde ederken kullandığı değer ölçütü ile Apriori algoritmasından farklılaşır. Predictive Apriori algoritması kuralların güvenilirliğini destek değerlerine göre değerlendiren doğruluk ölçütü ile kurallar üretir. Doğruluk değeri, hem destek hem de güven ölçütlerini kullanarak oluşturulan bir değerdir. Apriori algoritmasında, budama süreci, destek değerine göre yapılmakta ve sonrasında bulunan kurallar güvenilirlik derecesine göre kontrol edilmektedir. Tahminci Apriori algoritmasının kural oluşturma doğruluk değeri olduğu için, algoritma “n” adet en iyi birliktelik kuralını çıktı olarak verir. “n” rakamı kullanıcı tarafından belirlenen bir değerdir fakat kullanıcının verdiği “n” değerine göre kuralların da doğruluk oranları değişmektedir. Bu da verilen “n” değerine göre, Predictive Apriori kullanıldığında yüksek destek değerine sahip çok genel kurallar veya yüksek güvenilirlik, düşük desteğe sahip çok özgün kurallar üretebilmesinden kaynaklanmaktadır (Scheffer 2001).

4.5. Veritabanının Çözümlemesi

Amerikan Federal Havacılık Kurumu, havacılık kazaları ve olayları veritabanına ait örnek veri seti, 01/Ocak/2005 ve 31/Aralık/2009 tarihleri arasında elde edilmiş verileri içermektedir. Veriler, metin Belgesi (.txt) formatında paylaşılan 23,4 Megabyte büyüklüğünde, 15670 adet kayıttan oluşmaktadır. Bu verilere ait örnek görüntüler şekil 4.1.'de görülmektedir.

Verilere ait başlıca özellikler şunlardır:

- 15670 adet satırdan oluşmaktadır.
- 01.01.2005 Cumartesi – 31.12.2009 Perşembe tarihleri arasındaki uçak kaza ve olay bilgilerini içermektedir.
- 15670 adet uçak kaza ve olayına ait 164 adet özellik içermektedir.
- Veriler sayısal, metinsel, sayısal- metinsel ve tarih şeklinde bulunmaktadır.

4.6. Veritabanının Uygulama İçin Hazırlanması

Bir veri tabanının problemin tanımlanmasından sonra kullanılacak veri madenciliği programı için hazırlanması çalışma esnasındaki en zahmetli ve uzun süreçtir. Bu çalışmada verilerin hazırlanması için izlenen süreç aşağıda anlatılmıştır;

4.6.1. Verinin tabanın önışlemesi ve hazırlanması

Veri temizleme aşamasında, veri tabanından konu ile ilgisi olmayan bilgiler ayıklanmış ve eksik veriler tamamlanmıştır. Bu kapsamda; veri madenciliği çalışmasında yarar

sağlamayacağı değerlendirilen alanlar, veri tabanından çıkarılmıştır. Bu tip verilere örnek vermek gerekirse; aynı bilgileri içeren sütunlar olay tarih bilgisi, uçağın ve motorun grubu v.b. yinelenen veriler olduğu söylenebilir. Yine bu çalışma için gereksiz olan bilgiler; pist numarası, uçak kayıt, kaza araştırma ofisi bilgileri v.b. alanlar da veri tabanından çıkarılmıştır. Veri tabanından çıkarılan veriler arasında; verilerin alındığı form ve verileri araştıran ajans kodu, olay veya kazaya ait rapor türünü, olay veya kaza

c5	c1	c2	c3	c4	c6	c7	c8	c9	c10	c75	c132	c134	c136	c138	c139
20050101000019G	G		091	-23	F	2005	01	01	20050101		1215	9			
LAND-ROLL FW			MINR			OPDEF			EXEC			OTHR			UNK
20050101000199G	G		091	-23	F	2005	01	01	20050101		1430	9	B	A	
OMP MALF/FAIL	APPR-VFR	PTRN	CIRC		MINR			OPDEF			PERS			PLEA	
20050203001879G	G		091	-23	F	2005	02	03	20050203		1455	9	Z	Y	B
D LAND			LAND-LVL OFF	TOUCH	MINR			PILOT				INST			TRNG
20050203002299A	A		091	-23	F	2005	02	03	20050203		1455	9	B	E	
LAND-ROLL FW			SUBS			PILOT			INST				TRNG		
20050203002469A	A		091	-23	F	2005	02	03	20050203		1415	9			
TL	LAND-ROLL FW		SUBS			OPDEF					PERS			PLEA	
20050203002549A	A		091	-23	F	2005	02	03	20050203		1500	9	B	A	B
	APPR-FORCE LAND		SUBS			PILOT					INST			TRNG	
20050203034359A	A		091	-23	F	2005	02	03	20050203		1610	9	A	X	
	APPR-FINAL		DMOL			PILOT			PERS				PLEA		
20050204000649G	G		091	-23	F	2005	02	04	20050204		1400	9	B	C	
AND	LAND-LVL OFF	TOUCH	MINR			PILOT					INST			TRNG	
20050204001089A	A		091	-23	F	2005	02	04	20050204		1340	9	B	D	B
COLL GRD	CRUISE-FCD/PREC	LD	DMOL			PILOT					PERS			OTHR	
20050204001349C	C		121	-23	F	2005	02	04	20050204		1256	9	A	X	
	MINR		OPDEF			SCAC			UNKN				VFR		
20050204001519G	G		091	-23	F	2005	02	04	20050204		0715	9	B	A	
P LTS	GRND GRD TAXI		MINR			PILOT					PERS			PLEA	
20050204002439A	A		091	-23	F	2005	02	04	20050204		0920	9	A	X	
L GRD	CRUISE-NORMAL		DMOL			MISCL					EXEC			OTHR	
20050204004059A	A		091	-23	F	2005	02	04	20050204		1100	9	B	D	
	APPR-FINAL		SUBS			PILOT					PERS			PLEA	
20050204004389G	G		091	-23	F	2005	02	04	20050204		0315	9	B	E	
	LAND-LVL OFF	TOUCH	MINR			PILOT			PERS				PLEA		
20050204004759C	C		121	-23	F	2005	02	04	20050204		1250	9			
UISE-NORMAL			MINR			OPDEF			CRGO			ACAR			UNK
20050204007329A	A		137	-23	F	2005	02	04	20050204		1330	9	B	A	
ND-LVL OFF	TOUCH		SUBS			IMPRM			AGRI			AGRI			VFR
20050204021099G	G		091	-23	F	2005	02	04	20050204		1018	9	A	X	
CNTL	GRND GRD TAXI		MINR			PILOT					INST				TRNG
20050205000749G	G		091	-23	F	2005	02	05	20050205		0923	9	A	X	
MALF	DESCENT		NONE			OPDEF					INST			TRNG	
20050205000829A	A		091	-23	F	2005	02	05	20050205		1241	9	A	X	
IRIS POLES	APPR-FORCE LAND		DMOL			OPDEF					PERS			PLEA	

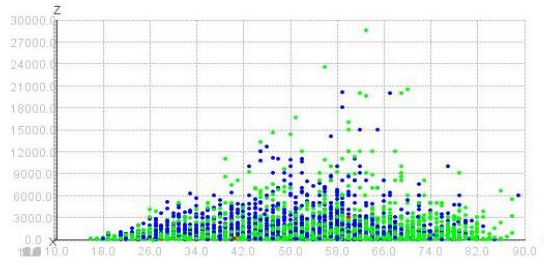
Şekil 4.1. Verilerin ham halinin gösterimi¹

araştırmasının ne boyutta yapıldığı, kazanın bildirilme şekli, sistem tarafından üretilen alanlar, uçağın ve motorun teknik sertifika veri sayfa bilgisi, uçak kaza ve olaylarının raporlanmasını sağlayan hazır formlar olan “form 19 ve form 4” ün ofise varış tarihleri, tamamlayıcı verinin ofise varış, kazaya ait dosyanın son değiştirilme tarihi, uçağın sahip ve işletmecisinin ismi, sahip veya işletmecinin eyaleti, hava operatörünün cihaz kodu, hava acentesinin sertifika numarası örnek olarak verilebilir.

¹ http://av-info.faa.gov/dd_sublevel.asp?Folder=\AID (30.05.2010)

4.6.2. Gürültülü ve aykırı veriler

Veri tabanında bulunan gürültülü veriler, tek tek sayıları kontrol edilip, çalışmaya etkileri sınanarak veri tabanından temizlenmiştir. Bu bağlamda, veri tabanında bulunan veri özelliklerinden; kumanda eden pilotun niteliği, sertifika türü, sınıf kategorisi, far parça numarası, genel kaza nedenleri, uçağın motor türü v.b.'nde bulunan gürültüler temizlenmiştir. Buna örnek vermek gerekirse; şekil 4.2.'de, genel kaza nedenlerine (C100) ait gürültüleri, Rapidminer programı ile gösterilmektedir.



Şekil 4.2. C100 verisine ait gürültünün gösterimi

4.6.3. Kayıp veriler

Kayıp veri nedeni ne olursa olsun istenmeyen bir durumdur. Kayıp verileri, bazı veri madenciliği teknikleri yok sayabilmektedir(karar ağaçları gibi.) Fakat lineer regresyon veya destek vektör makineleri gibi teknikler kayıp verilerle çalışamazlar. Ayrıca kayıp veriler ile çalışabilen algoritmalarda dahi tam veri girilmez ise tam veri girilerek çalıştırılan algoritmaya göre kesinlik konusunda eksiklik içerir.

Kayıp veriler için izlenecek ilk yol filtre etmektir. Bu yapıldığı zaman elimizdeki verileri %98 nispetinde kaybedebiliriz. Bu da çalışmayı doğru sonuçlara götürmez. Kayıp verilere ait özelliklerin yok edilmesi yine çalışmada eksikliklere neden olur; özellikle değerli özellikler kaldırılırsa çalışmanın anlamı ve doğruluğu kalmayabilir. Yine bu ve benzeri durumlar istenmeyen durumlardır.

Burada uygulanabilecek çözüm, kayıp verilere; ortalama, sıfır, sabit, en küçük veya en büyük veri değeri vermektir. Ya da çok ayrıntılı bir çalışma yapılmak isteniyorsa, özelliklere tek tek bakarak eksik değerler yok edilebilir. Fakat bu uygulamayı büyük veri tabanlarına uygulamak çok zaman ve uğraş gerektirir. Eksik verilerin yerine ortalama değer konduğu zaman, frekansı en yüksek veri, sonuç özelliğe yanıt olarak egemen

olabilir. Aynı şekilde eksik verilerin yerine, en az veya en çok deęerdeki veriler eklendiğinde, veriler en az veya en çok deęere yaklařabilir. Bu konuya dikkat edilmiřtir. Bilinen özelliklerin kayıp bölümlerine, eęer bu konuda kullanıcının yeteri kadar bilgisi varsa; istenilen deęer verilebilir. Örnek; Boeing 737 uçaęı ile uçmak için gerekli olan lisans tipi biliniyorsa ve bu uçakla kaza yapmış uçaęın lisans türü alanı kayıp ise bu alana bilinen lisans türü girilebilir. Bu uygulama; kayıp verilerin çok az ama çok önemli özellikleri için uygulanabilir.

Bu çalışmada veri tabanında bulunan kayıp veriler sayılarına ve ağırlıklarına göre deęerlendirilmiştir. %60'ı aşan kayıp veriye sahip özellikler modellerin performansını etkileyeceęi düşünülerek, veri tabanından çıkarılmıştır. Eksik veri sayısı az olan verilerde bilinen deęerler el ile girilmiştir.

Bunun dışındakiler SPSS 15.0 programının kayıp veri çözümleme (Missing Value Analysis) modülü kullanılarak tahmin edilmeye çalışılmıştır.¹

Veriler tahmin edilirken en iyisi elde edilmek için; beklenti en iyilemesi (expectation maximization) teknięi kullanılmıştır. Bu yöntem ile kayıp olmayan verilerden, matrisler yaratılır ve bu matrisler üzerinden elde edilen kovaryans matrisleri ile eksik veriler tahmin edilebilir. Bu teknięin kullanılma amacı;

- Güçlü ve üstün bir çözüm sağlamaı
- Tüm verileri kullanmasıdır.

Kayıp veriler, kayıp veri çözümleme modülü ile tahmin edildikten sonra, Microsoft Excel tablosu şeklinde kayıt edilerek, veri madencilięi programı için hazırlanmıştır.

4.6.4. Verinin dönüşümü

Veri madencilięinde kullanılan bazı teknikler yalnızca sayısal verilerle, bazıları da yalnızca metinsel veriler ile çalışır. Karar ağaęları her iki veri türü ile çalışabilir, fakat birliktelik kuralları yalnızca sözel verilerle çalışabilmektedir. Havacılık verileri; her ikisini de içerdieęi için bu çalışmada kullanılmaktadır. Ayrıca karar ağaęları algoritmalarının bazılarının, (Id3 gibi) yalnızca sözel verilerle çalışmasından dolayı, var olan verilerden sayısal olanlar ve çok çeřitli türde veri içeren özellikler kategorik olarak dönüřtürülmüřtür. Ayrıca ayrı ayrı olarak belirtilen, bazı veri alanları basit matematiksel işlemlerle toplanarak gruplandırılmıştır. Karar ağaęları kurallarında etiket özellięi, (hedef özellik) dokuz çeřit olan genel kaza nedenlerindeki insan hataları olarak seçilmiştir.

¹ <http://spss-for-windows.software.informer.com/15.0/> (10.12.2009)

Veri dönüştürülmesi aşamasında amaç; kaynak verinin, veri tipleri ve değerleri olarak değiştirilmesidir. Bu nedenle, yaş bölümünde doğum tarihi olan alan yaş olarak; pilotların yaşı alanı belli yaş aralıklarını belirten metin olarak, tarih alanındaki gün ve ay bilgileri metin olarak değiştirilerek girilmiştir. Ay bilgileri daha sonra mevsim değerlerine dönüştürülmüştür. Kazaların gerçekleştiği saat aralıkları, metin olarak kaydedilmiştir (Çizelge 4.2.).

Çizelge 4.2. Veri genel değişimi

Değişkenler	Kodlama şeması	Yeni değişkenler
Genel neden kategorisi	5,8,A= E Diğerleri=H	E,H
Sertifika türü	TT=FI, XX=FD, F9=FE 0=FK, 1=FA, 2=FY, F1=FB 3=FS, 9=FT, 99=FO, F3=FU	FI,FD,FK,FA,FY FB,FS,FT,FO,FU FE
Uçuş türü	1=K, 2=L, 3=M, 4=N 5=O, 6=P, 7=R, 8=S 9=T	K,L,M,N,O,P,R S,T
İkincil uçuş türü	1=AA,2=BB,3=CC	AA,BB,CC
Olay saati	0601-1200=S 1201-1800=O 1801-0600=A	S,O,A
Aylar	1,2,3=K, 4,5,6=I 7,8,9=Y, 10,11,12=S	K,I,Y,S
Uçak ağırlık sınıfı	1=U, 2=O	U,O
Uçak motor sayısı	1=T, Diğerleri=C	T,C
Uçak motor beygir gücü	3=U, 4=O	U,O
Olaya karışan uçak sayısı	1=F,2=S,3=T,9=N	F,S,T,N
Işık durumu	1=D, 2=N, 3=DA, 4=DN	D,DN,DA,DA
Yardımcı olay nedeni	1=AL, 2=AB,3=AC 4=AD, 5=AE, 6=AF 7=AG, 8=AI	AL,AB,AC,AD,AE AF,AG,AI
İkinci yardımcı olay nedeni	2=BC, 3=BD, 4=BE 5=BF, 7=BG, 8= BI	BC,BD,BE,BF,BG BI
Uçaktaki toplam yolcu sayısı	0=Y, 1-10= A 11-100=N, 101-500=C	Y,A,N,C
Toplam ölen ve yaralanan	0=Y, 1-10= A 11-60=N	Y,A,N
Yaralan yer personeli	1=V, 0=Y	V,Y
Pilotların yaşları	34 ve altı=G, 35-42=O, 43-50=E 51 ve üstü= Y	G,O,E,Y
Federal havacılık kuralları numarası	91=G, 103=U, 105=P 121=O, 125=C, 133=H 135=T, 137=A,141=F	G,U,P,O,C,H,T,A,F
Günler	1-7=AI, 8-15=AK 16-23=AC, 24-31=AD	AI,AK,AC,AD
Uçak gövde uçuş saati	0-499=B, 500-999=N 1000-1499=E, ... ,50000 ve üzeri= BG	B,N,E,...,BG
Pilot sınıf kategorisi	O=A, 1=B, 2=C, 5=D 6=E, ...,70=QX	A,B,C,D,E,...,QX

4.6.5. Veri ayırıştırma

Bu çalışmada elde edilen veriler; “5M yöntemi” dikkate alınarak kategorilere ayrılmıştır. Bu amaçla makine bölümünde; uçak ile ilgili olan tasarım, üretim, uçak

gövdesi ve bakım ile ilgili özellikler, görev bölümünde; uçuşun operasyon bilgisi ile ilgili özellikler, ortam bölümünde iki türlü özellik dikkate alınarak; doğal ortamda hava, arazi yapısı, ısı, ışık gibi özellikler; yapay ortamda insan yapısı özellikler; havaalanı, pist özellikleri, insan bölümünde; yaş, uçuş saati, pilotun mesleği ile ilgili özellikler dikkate alınarak veri tabanında bulunan veriler ayrıştırılmıştır.

Bunun haricinde ayrıştırma esnasında uçak kaza ve olayı sonrası, elde edilen bilgilerde bir grup haline getirilmiştir. Bu özellikler arasında uçağın hasar durumu, kaza türü, ölen ve yaralan personel sayısı v.b. bilgiler yer almıştır.

Veri temizlenmesi ve dönüştürülmesinden sonraki veri tabanının son hali çizelge 4.3.'te sunulmuştur.

Çizelge 4.3. Verinin temizlenmesi ve dönüştürülmesinden sonra veri tabanının son hali

Sütun	Sütun türü	Uzunluk	Tanım
c8	Metin	2	Kaza veya olayın olduğu gün
c10	Metin	4	Kaza veya olayın olduğu lokal zaman
c7	Metin	2	Kaza veya olayın olduğu ay
c6	Metin	4	Kaza veya olayın olduğu yıl
c5	Sayı	5	AID ana tablosu ile ilişkilendirilen kontrol numarası
c15	Metin	30	Kaza veya olayın gerçekleştiği havaalanı
c31	Metin	5	Uçağın gövde uçuş saati
c27	Metin	2	Uçağın bölge sertifikası
c102	Metin	1	Birinci derece uçuş türü
c104	Metin	1	İkinci derece uçuş türü
c233	Metin	1	Pilotun yaşam durumu
c145	Metin	1	Uçağın ağırlık sınıfı
c153	Metin	1	Uçak motorunun beygir gücü
c151	Metin	2	Uçağın motor sayısı
c157	Metin	2	Uçağın iniş takım türü
c155	Metin	1	Uçağın motor türü
c49	Metin	1	Kumanda eden pilot niteliği
c52	Metin	1	Kumanda eden pilotun uzmanlığı
c47	Metin	1	Kumanda eden pilotun başka herhangi kategori sınıfı
c53	Metin	5	Pilotun o marka ve model uçakta son 90 gün uçuş saati
c55	Metin	5	Pilotun son 90 gün içerisindeki uçuş saati
c50	Metin	2	Kumanda eden pilotun yaşı
c78	Metin	2	Birincil olay ve kaza nedeni
c80	Metin	2	Yardımcı olay ve kaza nedeni
c98	Metin	1	Uçağın hasar kodu
c94	Metin	2	Kaza kodu türü
c82	Metin	1	Kazada yer alan insan sınıfı
c96	Metin	2	Uçuş pozisyonu türü
c100	Metin	1	Genel kaza ve olay neden kategorisi
c110	Metin	1	Işık durumu
c117	Metin	4	Pist durumu
c71	Metin	2	Yaralanan yer personeli sayısı
c72	Metin	2	Ölen yer personeli sayısı
c61	Metin	3	Toplam Personel ve yolcu sayısı
c106	Metin	1	Birinci derece uçuş durumu
c108	Metin	1	İkinci derece uçuş durumu
c45	Metin	2	Kumanda eden pilotun sınıf kategorisi
c41	Metin	2	Kumanda eden pilotun sertifika türü
c2	Metin	1	Federal havacılık kuralları numarası

4.7. Veri Madenciliği Yöntemlerinin Uygulanması

Bu çalışma da öncelikli olarak seçilen veri madenciliği yöntemlerinden sınıflandırma ve birliktelik kuralı kullanılacaktır. Sınıflandırmanın tercih edilmesi tahminleyici bir model olması ve geleceğe yönelik çıkarımlar elde edebilmesidir. Birliktelik kuralları ise veriler arasındaki ilişkilerin tespit edilmesi ve eş zamanlı olarak gerçekleşme olasılıklarını ortaya koyarak çözümleme yapmaktadır. Her iki yönteminde uygulaması ve modellerin oluşturulması aşağıda gösterilmiştir;

4.7.1. Sınıflandırma

Sınıflandırma yöntemi veri madenciliği için kullanılanlar arasında çok önemli bir yere sahiptir. Bölüm 3'te bahsedilen birçok sınıflandırma yöntemleri arasında en çok kullanılanı karar ağaçlarıdır. Bu çalışmada Rapidminer programı içerisinde yer alan karar ağacı algoritmalarından dört adedi denenerek modeller oluşturulmuştur. Hazırlanan veriler algoritmaların kullanımına uygun duruma getirildikten sonra, seçilen algoritmaların ne kadar doğru sonuçlar verdiğini test etmek amacı ile doğrulama yöntemlerine başvurulmuştur. Doğrulama yöntemleri de doğrulama yeteneklerine göre sınanmıştır.

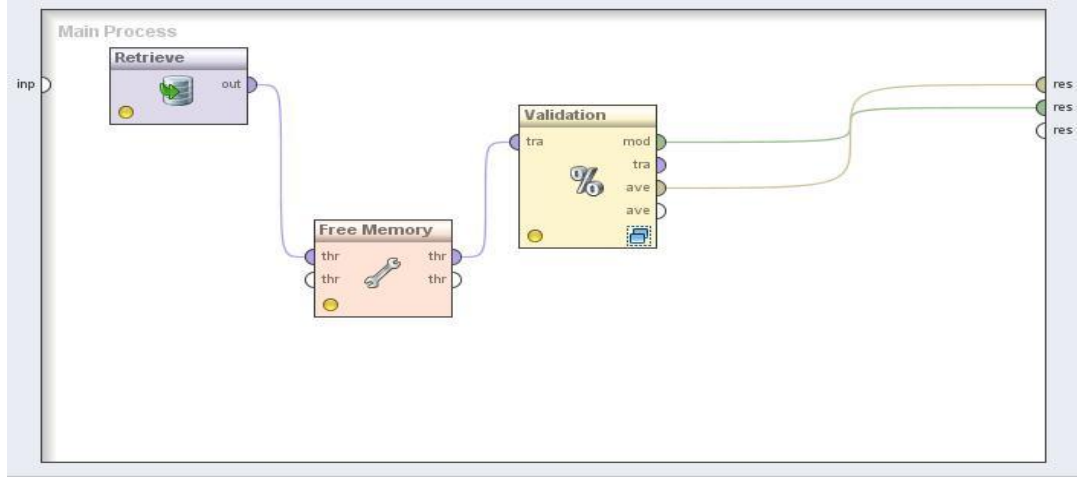
a) Modelin kurulması ve doğrulanması

Doğrulama yöntemleri tüm veriyi kullanmayı sağlayan bir metottur. Veri seti doğrulama yönteminin çeşidine göre iki kısma ayrılır. İlk veri setiyle; model kurulur, kurulan model diğer veri setiyle test edilir ve doğruluk oranı hesaplanır. İkinci veri setiyle ise model kurulur ve kurulan model birinci veri setiyle tahmin edilip, doğruluk oranı hesaplanır. En sonunda tüm veriler kullanılarak model oluşturulur. Daha önceden hesaplanmış olan iki doğruluk oranının ortalamasıyla, son model karşılaştırılır.

Bu kapsamda kullanılan programın sahip olduğu doğrulama yöntemlerinden sınıflandırma özelliğine uygun olarak üç tanesi (Split validation, X-validation, Bootstrapping validation) denenmiştir. Elde edilen sonuçlardan çıkan test verileri değerlendirildiğinde; özellikle doğruluk oranına göre, çalışma hızı ve verdiği net değerlerden dolayı Split validation doğrulama yönteminin bu çalışma için uygun olduğuna karar verilmiştir.

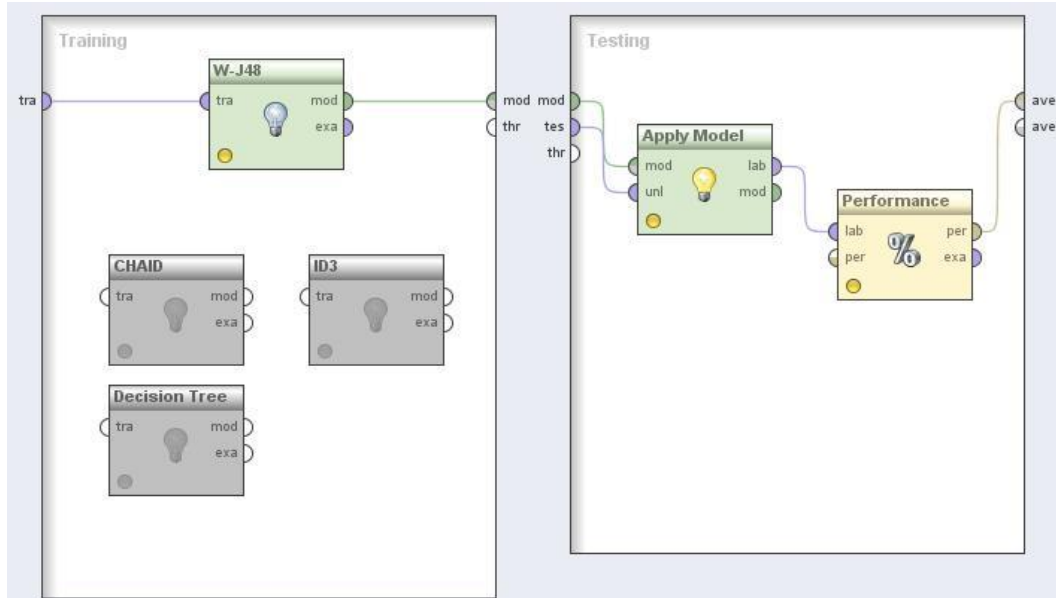
Split validate aracı, temel olarak eldeki veriyi eğitim verisi ve test verisi olmak üzere ikiye bölerek oluşturulan modelin performansını değerlendirmektedir (Şekil 4.3.).

Verileri eğitim ve test verisi olarak ayırırken göreceli yaklaşım kullanıldığından, model kurulurken tüm verilerin %70'i eğitim verisi, % 30'u da test verisi olarak değerlendirilmiştir.



Şekil 4.3. Split validate doğrulama aracının veri ile ilişkilendirilmesi

Doğrulama yöntemleri test edilirken, bu çalışmada kullanılacak olan algoritmalar ayrı ayrı denenmiş ve en doğru sonucun elde edilmesi hedeflenmiştir. Çalışmada kullanılan verilere dört farklı mantık ile çalışan karar ağacı algoritmaları sırasıyla uygulanmıştır.(Şekil 4.4.). Uygulama sonucunda elde edilen çizelgeler çizelge 4.4. ile 4.7. arasında gösterilmiştir.



Şekil 4.4. Split validate doğrulama aracının gösterimi

Uygulanan algoritmalarından elde edilen başarımlar karşılaştırıldığında:

- Doğruluk değerinin büyüklüğü, kurulan modelin sınıflandırma hatasının bir o kadar düşük olduğunu ve başarı oranının yüksek olduğunu gösterir.
- Kappa istatistiği değeri, genel olarak 1 ile 0 arasında değişmektedir. Değer 1'e ne kadar yaklaşırsa veriler birbirine o kadar uyumaktadır. Elde edilen değer 0'a yaklaştıkça uyuma yalnızca şans ile belirlenebilir ve diğer hallerde hiç bir uyuma yoktur.

Çizelge 4.4. W-J48 algoritmasının başarı ölçütleri

Algoritma	W-J48				
Başarım ölçütleri	Accuracy %	Kappa	Weighted mean recall%	Weighted mean precision %	Correlation
Görev	71,04	0,318	64,50	69,96	0,340
İnsan	73,03	0,157	56,19	66,26	0,201
Kaza	96,90	0,813	86,96	95,39	0,819
Makine	73,81	0,142	55,43	66,79	0,191
Ortam 1	72,79	0,097	53,61	68,93	0,165
Ortam 2	73,22	0,159	56,24	66,84	0,205

Çizelge 4.5. CHAID algoritmasının başarı ölçütleri

Algoritma	CHAID				
Başarım ölçütleri	Accuracy %	Kappa	Weighted mean recall%	Weighted mean precision%	correlation
Görev	65,03	0,159	57,11	61,07	0,177
İnsan	71,63	0,098	53,81	61,57	0,133
Kaza	92,74	0,475	68,06	87,09	0,518
Makine	72,21	0,122	54,84	61,30	0,148
Ortam 1	71,84	0,013	50,46	79,23	0,073
Ortam 2	72,43	0,141	55,58	64,16	0,178

Çizelge 4.6 ID3 algoritmasının başarı ölçütleri

Algoritma	ID3				
Başarım ölçütleri	Accuracy %	Kappa	Weighted mean recall%	Weighted mean precision%	correlation
Görev	63,72	0,215	60,68	60,81	0,215
İnsan	67,41	0,170	58,20	58,84	0,170
Kaza	96,24	0,773	85,21	93,10	0,779
Makine	71,72	0,136	55,58	60,66	0,154
Ortam 1	72,75	0,096	53,56	68,66	0,163
Ortam 2	59,73	0,113	56,33	55,38	0,117

Çizelge 4.7. Desicion tree algoritmasının başarı ölçütleri

Algoritma	Desicion Tree				
Başarım ölçütleri	Accuracy %	Kappa	Weighted mean recall%	Weighted mean precision%	correlation
Görev	63,27	0,000	50,00	31,63	Unknown
İnsan	71,63	0,007	50,21	59,40	0,028
Kaza	93,97	0,565	71,51	93,92	0,615
Makine	72,62	0,000	50,00	36,13	Unknown
Ortam 1	71,60	0,000	50,00	35,80	Unknown
Ortam 2	71,71	0,000	50,00	35,85	Unknown

- Ağırlıklı ortalama değerleri yapılan sınıflandırmanın, ağırlıklı ortalama değerini ve kesinliğini göstermektedir.
- Korelasyon değeri, veri dizisi için belirlenen etiket değeri ve tahmin etme yeteneği arasındaki ilişkiyi göstermektedir.

Elde edilen değerlerin tümünün doğruluk oranı ile doğru orantılı olduğu gözlemlenmiştir. Özellikle doğruluk değeri olarak algoritma ile oluşturulan modelin doğruluğunu test eden accuracy değerine göre W-J48 algoritmasının daha doğru sonuçlar verdiği görüldüğünden bu algoritmanın sonuçları uçak kaza ve olayları için değerlendirilmiştir.

b) Veri gruplarının değerlendirilmesi

- İnsan:
 - Aletle uçuş sertifikasına sahip pilotun kumanda ettiği, pilotun ölmediği, pilotun meslek olarak bilinmediği durumlarda, aynı koşullar geçerli olmak kaydıyla: Pilotun profesyonel pilot olduğu, son 90 gün içerisindeki toplam uçuş saati 0 ila 100 saat arasında olan ve pilotun yaşının 43-50 arası değiştiği durumlarda, son 90 gün içerisindeki toplam uçuş saati 250 ila 500, 750 ve üzeri olduğu durumlar,
 - Aletle uçuş sertifikasına sahip pilotun kumanda ettiği, pilotun öldüğü, pilotun kategorisinin planör, helikopter ve uçuş öğrencisi, bilinmediği durumlarda,
 - Aynı koşullar geçerli olmak kaydıyla pilotun kategorisinin başka ülke tarafından sınıflandırıldığı, pilotun yaşının 42 ve altı olduğu durumlarda insan hatası gözlemlenmektedir.
- Görev:
 - Aletli uçuş kurallarının uygulandığı, genel operasyon ve havacılık kurallarının geçerli olduğu, uçuş pozisyonun tırmanış, kalkış ve yaklaşma/iniş durumunda iken,
 - Aynı başlangıç koşullarında, uçuş pozisyonun alçalma olduğu 2007 ve 2008 yıllarında genel olarak, 2006 yılında, 12:00 ila 18:00 saatleri arasında gerçekleşen uçuşlarda meydana gelen kaza ve olaylarda,
 - Aletli uçuş kurallarının uygulandığı, talep üzerine gerçekleşen ve yönetici seviyesinde devlet görevlileri için gerçekleşen uçuşlarda meydana gelen kaza ve olaylarda,
 - Görerek ve özel görerek uçuş (hava alanı trafiğine girilen noktalar gibi özel kontrol bölgelerinde gerçekleşen uçuşlar.) koşullarının uygulandığı uçuşlarda meydana gelen kaza ve olaylarda genel olarak,
 - Öğrenci pilotların radar kontrollü olarak yaptığı görerek uçuşlardan 06:00 ila 12:00 saatleri arasında gerçekleşen uçuşlarda, yine kaza ve olayların,
 - Uçuş plan tanımının bilinmediği, genel operasyon ve havacılık kurallarının geçerli olduğu, iş amaçlı yapılan fakat özel amacın bilinmediği uçuşlarda, uçuş pozisyonun yaklaşma/iniş, uçuş saatlerinin 12:00 ila 18:00 olduğu durumda insan hataları gözlenmiştir.

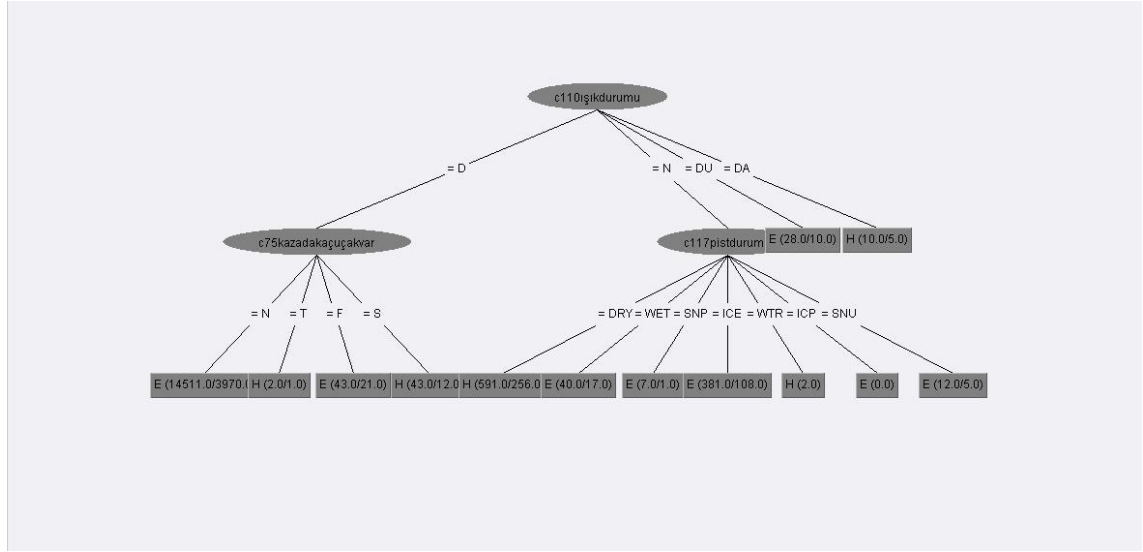
Bu veri grubuna ait elde edilen diğer karar ağacı kuralları EK 1’de sunulmuştur.

- Kaza:
 - Kaza ile ilişkili insan olarak yalnızca pilotların olduğu, yanlış hükümden dolayı kaza ve olayın gerçekleştiği uçuşlar da,
 - Kaza ile ilişkili insan olarak yalnızca pilotların (neden) olduğu, ihmalkârlıktan dolayı kaza ve olayın gerçekleştiği uçuşlar da, uçakta bulunan uçuş personeli ve yolcunun 11 ila 100 arasında olduğu, kazaya katkısı olan nedenler arasında hava, buzlu/çamurlu pist ve birkaç tanesi bir arada olduğu durumlarda,
 - Aynı koşullar geçerli olmak kaydıyla, uçakta bulunan uçuş personeli ve yolcunun olmadığı, 1 ila 10 arasında olduğu durumlarda,
 - Kaza ile ilişkili insan olarak yalnızca pilotların olduğu, dikkatsizlik sonucu gerçekleşen kaza ve olayın olduğu uçuşlar da,
 - Kaza ile ilişkili insan olarak yalnızca pilotların olduğu, pilotun vasıfsız olduğu için meydana gelen kaza ve olaylarda, kazaya katkısı olan nedenler arasında hava, buzlu/çamurlu, ıslak pist, birkaç tanesi bir arada, suyun kristalleşmeye başladığı sıcaklık, yabancı madde hasarı, yoğunluk yüksekliği, denize mecburi iniş, yüksek bitki örtüsü, düşük banket, acil iniş ve planlanmamış başka piste iniş vb. olduğu durumlarda,
 - Aynı koşullar geçerli olmak kaydıyla, kazaya katkısı olan nedenler arasında aşağı doğru hava hareketi olduğu zaman, uçakta meydana gelen hasar kodu nerdeyse tamamen ve yıkım derecesinde olduğu durumda,
 - Kaza ile ilişkili insan olarak uçak mühendisinin olduğu, toplam ölen insan sayısının 1 ila 10 olduğu zaman,
 - Kaza ile ilişkili insan olarak kontrol pilotu ve sertifika almamış öğrenci pilotun olduğu uçuşlarda; uçakta bulunan uçuş personeli ve yolcunun olmadığı, 1 ila 10 arasında olduğu durumlarda insan hatası gözlemlenmiştir.

Bu veri grubuna ait elde edilen diğer karar ağacı kuralları EK 1 'de sunulmuştur.

- Makine:
 - Uçak ağırlığının 12500 libreden fazla olduğu, iniş takım türünün 3 tekerlekli ve uçuş esnasında iniş takım muhafazasına geri giren sisteme sahip olan, turbo aks motor türüne sahip, Avrupa ve kuzeybatı dağlar bölgesi motor sertifikasına sahip uçaklarda,
 - Yine uçak ağırlığının 12500 libreden fazla olduğu, iniş takım türünün üç tekerlekli ve iniş esnasında iniş takım muhafazasına geri giren sisteme sahip olan, piston V motor türüne sahip, güneybatı ve kuzeybatı dağlar bölgesi motor sertifikasına sahip uçaklarda,
 - Bundan başka uçak ağırlığının 12500 libreden fazla olduğu, iniş takım türünün üç tekerlekli ve iniş esnasında iniş takım muhafazasına geri giren sisteme sahip olan, jet motor türüne sahip uçaklarda meydana gelen kaza ve olaylarda,
 - Uçak ağırlığının 12500 libreden fazla olduğu fakat iniş takım türünün konvansiyonel ve uçuş esnasında iniş takım muhafazasına geri giren sisteme sahip olan uçaklarda,
 - Uçak ağırlığının 12500 libreden az olduğu uçaklarda genel olarak meydana gelen kaza ve olaylarda insan hatası gözlenmiştir.

- Ortam 1:
 - Uçuştaki ışık koşulları gündüz, gerçekleşen kaza ve olaya karışan uçak sayısı bir olduğu durumlarda, kazanın havada uçuş esnasında değil de, yerde veya havaalanında olduğu durumlarda,
 - Uçuştaki ışık koşullarının gece, pistin durumunun ıslak, tamamen karlı, kardan temizlenmiş, buzlu ve kısmen buzlu olduğu durumlarda,
 - Bundan başka ışık durumunun alacak karanlık olduğu durumda meydana gelen kaza ve olaylar da insan hatasının olası olduğu gözlemlenmiştir (Şekil 4.5.)



Şekil 4.5. Ortam 1 veri grubunun karar ağacı görüntüsü

- Ortam 2:
 - Uçuştaki hava koşullarının donma noktasına yakın, rüzgarlı, hafif yağmur, sisli, kasırga, gök gürültülü fırtına, hafif yağmurlu, hafif ve kuvvetli kar yağışı koşullarında, bulut alt tavanın düşük olduğu ve dağ rüzgarının etkili olduğu durumlarda,
 - Yine kuvvetli yağmur altında aletle yapılan uçuşlarda insan hatası gözlenmiş, havanın bir neden olarak değerlendirilmediği ve görerek koşullarda yapılan uçuşlarda insan hatası sonucu kaza ve olayların gerçekleşmesi muhtemel olarak,

4.7.2. Birliktelik kuralı

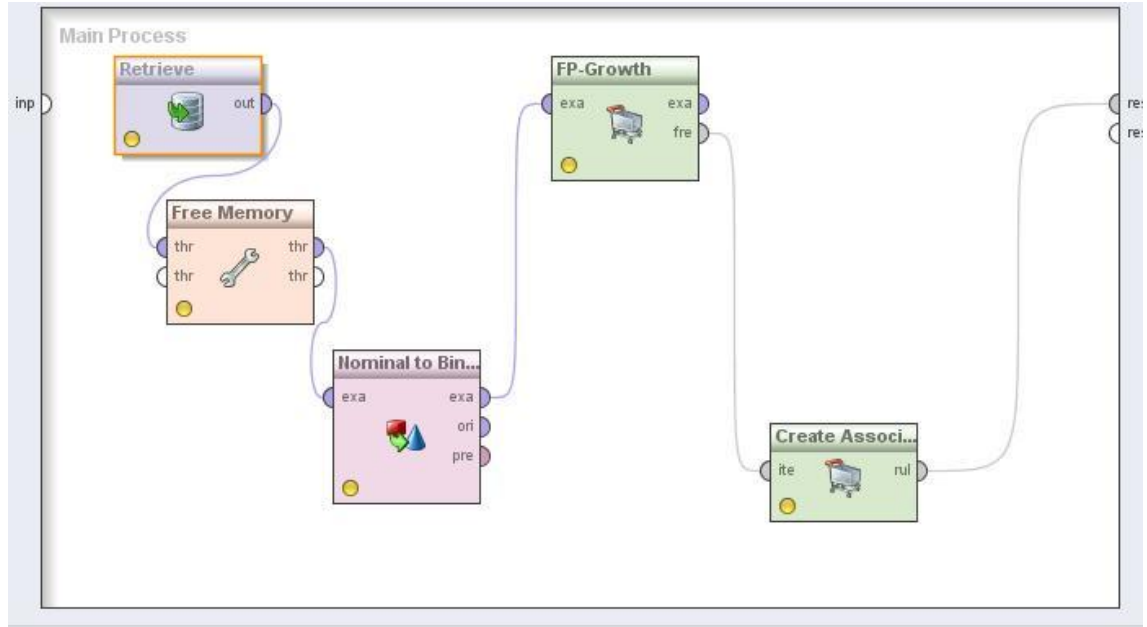
Bu çalışmada birliktelik kurallarının elde edilmesi için Rapidminer programının içerdiği algoritmalar ile modeller oluşturulmuştur.

a) Modelin kurulması ve değerlendirilmesi

Önişleme ile hazırlanan verilere üç adet birliktelik algoritması uygulanmıştır. Algoritmalar uygulanırken;

- Fp-growth Algoritması için en düşük destek ve ölçütü olarak 0,95 alınmıştır (Şekil 4.6.).
- W-apriori algoritması için en düşük destek ölçütü 0,1, en düşük güven ölçütü olarak da 0,95 alınmıştır.
- W-Predictive apriori algoritması güven ve destek ölçütü yerine destek ölçütünü güven ölçütü ile kıyaslayarak oluşturulan doğruluk oranı kullanıldığı için özel olarak destek ve güven ölçütü değerleri verilememiştir. Fakat doğruluk ölçütüne etki eden 'n' kural sayısı 10 olarak alınmıştır.

Elde edilen kuralların ortalaması alınırken, üretilen kural sayısı 10'dan büyük ise yalnız 10 kuralın, altında ise elde edilen kural kadar ortalama alınarak çizelgeye eklenmiştir.



Şekil 4.6. Fp-growth algoritmasının veri ile ilişkilendirilmesi

Elde edilen sonuçlar karşılaştırılarak çizelge 4.8.'de oluşturulmuş ve en uygun algoritma ile elde edilen birliktelik kuralları çizelge 4.9.'da, uçak kaza ve olayları için değerlendirilmiştir.

Çizelge 4.8. Birliktelik kuralı için kullanılan algoritmaların karşılaştırılması

Algoritma	Fp-growth	W-Predictive Apriori	W-apriori
	Ort. güven	Ort. doğruluk	Ort. güven
Görev	0,9998	0,99385	0,98
İnsan	0,9971	0,99328	0,99
Kaza	0,99365	0,9966	0,9925
Makine	0,99484	1	Kural üretilmemiştir.
Ortam 1	0,56345	0,9714	Kural üretilmemiştir.
Ortam 2	0,89449	0,99833	Kural üretilmemiştir.

Çizelge 4.9. Birliktelik kurallarının elde edilmesi için seçilen algoritma

Algoritma	Seçilen algoritma
Görev	Fp-growth
İnsan	Fp-growth
Kaza	W-predictive apriori
Makine	W-predictive apriori
Ortam 1	W-predictive apriori
Ortam 2	W-predictive apriori

Performanslarına göre değerlendirilerek seçilen algoritmalarından elde edilen sonuçlar değerlendirildiğinde, birliktelik kuralı ile elde edilen kuralların çoğunda insan hatasının sınıflandırıcı olarak var olmadığı görülmektedir (Şekil 4.7.). Bunun nedeninin de birliktelik kurallarının bütün özellikleri, sınıflandırıcı nitelik olarak değerlendirmesinden kaynaklanmaktadır. Bu yüzden birliktelik kuralı ile elde edilen değerlerin sonuçları değerlendirilmemiş EK 2’de yayınlanmıştır.

W-PredictiveApriori

PredictiveApriori

=====

Best rules found:

1. c31gövdeuçuşsaati=BA c157iniştakım=RC C27BÖLGESERTİFİKA=SO 275 ==> c151motorsayısı=C 275 acc:(0.99495)
2. c157iniştakım=RC C155MOTORKOD=R C27BÖLGESERTİFİKA=SO 257 ==> c151motorsayısı=C 257 acc:(0.99493)
3. c145ağırlık=OV c100insanhatası=H C27BÖLGESERTİFİKA=NM 206 ==> c153beygörgücü=0 c151motorsayısı=C 206 acc:(0.99487)
4. c145ağırlık=OV c153beygörgücü=0 c100insanhatası=H C27BÖLGESERTİFİKA=NM 206 ==> c157iniştakım=RT 206 acc:(0.99487)
5. c157iniştakım=RT C27BÖLGESERTİFİKA=EA 195 ==> c151motorsayısı=C 195 acc:(0.99485)
6. c145ağırlık=OV c31gövdeuçuşsaati=BC 181 ==> c153beygörgücü=0 181 acc:(0.99481)
7. c145ağırlık=OV C27BÖLGESERTİFİKA=EA 176 ==> c153beygörgücü=0 c151motorsayısı=C 176 acc:(0.9948)
8. c31gövdeuçuşsaati=BD C27BÖLGESERTİFİKA=NM 171 ==> c145ağırlık=OV c153beygörgücü=0 171 acc:(0.99478)
9. c31gövdeuçuşsaati=BD C27BÖLGESERTİFİKA=NM 171 ==> c145ağırlık=OV c151motorsayısı=C 171 acc:(0.99478)
10. c145ağırlık=OV c31gövdeuçuşsaati=BD C155MOTORKOD=T 169 ==> c151motorsayısı=C 169 acc:(0.99477)

Şekil 4.7. Makine veri grubuna ait birliktelik kuralları

4.7.3. Birliktelik kuralı ve karar ağaçları arasındaki farklar

Her iki veri madenciliği tekniği de sınıflandırma yapabilmektedir fakat bu iki yöntem arasındaki fark/farklar nedir/nelerdir? Bu çalışmada elde edilen kurallar ve sonuçlara göre dört adet fark aşağıda sıralanmıştır:

- Birliktelik kuralları yalnızca sözel veri grupları ile çalışabilmektedir, bu yüzden sayısal verilerin veri tabanından uzaklaştırılması veya dönüştürülmesi gerekir fakat diğer yandan da karar ağaçları her ne kadar sayısal verilerle çalışabilen algoritmalara sahip olsa dahi sözel veriler ile daha yüksek doğruluk oranı verdiği gözlemlenmiştir.
- Karar ağaçlarının aksine birliktelik kuralları, kural üretme aşamasında bütün kuralları göz önünde bulundurmaz, özelliklerin alt-dizilerinin kombinasyonları üzerinde çalışır. Bu çalışmada da elde edilen kurallarda; iki elemanlı, üç elemanlı hatta dört elemanlı kurallar üretilmiştir.
- Başka bir anlamda karar ağaçları ve birliktelik kuralları birbirine benzemektedir. Fakat aradaki en büyük fark; karar ağaçlarının yalnızca bir özelliği etiket olarak belirlemesi ve bu etiket özelliği sınıflandırıcı olarak kullanmasıdır. Birliktelik kuralları ise her özelliği ve her özellik grubunu sınıflandırıcı olarak kullanması, daha sonra da oluşan her birleşimde sınıflandırma yapmasıdır. Bu çalışmada da karar ağaçlarında insan hatası sınıflandırıcı özellik olarak ele alınmış ve insan hatası ile sınıflanmış kurallar elde edilmiştir. Birliktelik kurallarında tam olarak böyle bir durum söz konusu değildir.
- Birliktelik kuralları üretilirken, üretilen kuralların değerlendirilmesinde güvenilirlik ve destek değerleri ile sınırlamalar getirilebilir. Predictive apriori algoritmasında olduğu gibi kuralları “n” tane ile sınırlandırmak olasıdır; fakat karar ağaçlarında her ne kadar budama ve ön budama teknikleri kullanılsa da elde edilen kuralları eğer belirli bir özellik üzerine yönelme yapılmadıysa sınırlamak olası değildir.

5. SONUÇ ve DEĞERLENDİRME

5.1. Genel sonuçlar

Havacılık endüstrisi, her anlamda gelişen ve ilerleyen bir alandır. Bu bağlamda Türk havacılığı da, uluslararası kürsü de saygın bir yer edinmek ve maddi manevi kayıpları önlemek amacı ile uçuş faaliyetlerinde, gelişen teknolojik yöntemler kullanmalıdır. Bu sayede ekonomik anlamda kazançlar da elde edilmiş olacaktır. Fakat bu gelişme ne kadar ileri olursa olsun, uçak kazaları meydana gelmektedir. Kazaların önlenmesi için, havacılık güvenliğinin ve insan makine ilişkisinin en son ve en zayıf halkası olan insan etmeninin her zaman kuvvetli tutulması gerekir. Bunu yapabilmek için de bu etmenin, hataya dönüşmeden önce olası tüm öncülerinin fark edilip engellenmesi gerekmektedir. Bunu yaparken de tüm uçak kazaları önlenemeyecek olursa da azaltılması için en iyi yollardan biri, daha önce gerçekleşen kaza ve olaylardan dersler çıkartılmasıdır.

Türkiye’ de sivil ve askeri uçuşların, dolayısıyla iç hava trafiğinin hızla artacağı öngörülmektedir. Özellikle Diğer Euro Kontrol üyesi ülkelerin hepsinde neredeyse %6 ile %8 oranında bir azalma yaşanırken, Türk sivil havacılığında %3'lük bir artış beklenmektedir¹. Türkiye'nin bulunduğu coğrafi konum ve havacılık alanındaki gelişmeler buna neden olan önemli etmenlerdir.

Yine Türk Hava Kuvvetleri'nin, 2015 yılına kadar bünyesine katacağı 10 adet A-400M nakliye uçağı, 2012 yılına kadar alacağı 55 adet K-T1 eğitim uçağı, 30 adet F-16 Blok 50 savaş uçağı, projenin tamamlanmasına müteakip alacağı dört adet havadan kontrol ve ihbar uçağı ile sayısı proje bitiminde teslim alacağı Joint Strike Fighter (JSF) savaş uçakları ile hava trafiğinin artacağı tahmin edilmektedir.²

Bu tez çalışması, ulaştırma kazaları içerisinde sayı olarak küçük fakat yarattığı sonuçlar ile büyük yer kaplayan hava kaza ve olaylarını incelemektedir. Çalışmada veri madenciliği uygulamalarında öncelikli olarak seçilen sınıflandırma ve birliktelik kuralları yöntem olarak kullanılmıştır. Çalışmanın amacına uygun olarak; ikinci bölümde insan etmeni ve hatası ile havacılık güvenliği konusu, üçüncü bölümde veri madenciliği yöntemi, dördüncü bölümde uçak ve kaza olay veri kümesine Rapidminer 5.0.6 programı ile karar ağacı ve birliktelik kuralı algoritmaları uygulanmış, sonuçları verilmiştir.

¹ http://www.aimsa.org.tr/index.php?option=com_content&view=article&id=30%3Aeurocontrol-icinde-havacilikta-tek-gelime-goesteren-uelke-tuerkiye&catid=3%3Ahavacilik-haber&lang=tr (25.05.2010)

² <http://www.hvkk.tsk.tr/PageSub/Calismalarimiz/Projeler/F-16Blok50.aspx> (25.05.2010)

Veri madenciliği ile verilerin çözümlenmesi son yıllarda bütün alanlarda uygulanan bir yöntem olarak karşımıza çıkmaktadır. Özellikle ticari hayata önemli yenilikler ve kolaylıklar getiren bir disiplindir. Her gün farklı şekillerde depolanan verilerin, depolama araçlarının ucuzlamasıyla artması ve bu verilerin değerlendirilmesi gerekliliğine olan inanç, bugün bütün alanları veri madenciliğine yönlendirmiştir.

Bu şekilde havacılık alanında dar boğaz yaratan en önemli konu kazalar ve etkisi seçilmiştir. Bu seçimden hareketle en büyük etmen olan insan etmeni, çözümü ve ilişkileri ile ilgili olarak geleceğe yönelik karar verilmesini kolaylaştıracak kurallar çıkarılmaya çalışılmıştır.

Çalışmada elde edilen sonuçlar değerlendirildiğinde, veri madenciliği sayesinde kesin kuralların elde edildiği görülmüştür. Elde edilen bu kurallar çerçevesinde insan hatasının oluşmasını sağlayan birçok nedenin hangi noktalarda kesiştiği ve hataya yol açtığı gözler önüne serilmiştir.

5.1.1. Sonuçların irdelenmesi

Veri madenciliği sayesinde elde edilen sonuçlar doğrudan bir program ve bilgisayar tarafından yapılmıştır. Bu değerlendirme havacılık alanı için bir karar sonucu değil, karar sürecine etki edecek bir bulgudur. Genel olarak elde edilen kurallar mantık çerçevesinde bakıldığında doğrudur ve ilgi çekicidir. Geliştirilebilen sistem, esnek bir yapıya sahip olduğundan, sağlam veri dizisi ile desteklenecek her özellik tezde izlenen aşamalar ile incelenebilir ve sistem geliştirilebilir. Gerçek kazalardan elde edilen bu veriler, farklı veri madenciliği algoritmalarının uygulanmasına olanak tanıyacak yapıdadır.

Veri boyutunun ve özelliklerinin çok olması, sorgu zamanlarının uzun olmasına neden olmuştur. Daha yüksek özelliklere sahip bilgisayar donanımları ile kullanılan algoritmalar daha hızlı ve daha çok denenebilir özelliklerde olacaktır. Farklı ölçütlerin algoritmalar üzerinde testleri daha kolay olabilir.

5.1.2. Yöntemin irdelenmesi;

Veri madenciliğinde seçilecek yöntem önemlidir. Seçilen yönteme göre elde edilen sonuçlar çok farklı olabilir. Veri madenciliğinde en önemli olan verinin sağlıklı ve doğru olmasıdır, yanlış verilerle elde edilecek sonuçların da yanıltıcı olması tabiidir.

Her veri grubu; başlı başına bir farklılık içerdiği için, oluşturulan kurallar da çok çeşitli ve dallanmıştır. Örnek olarak; incelenen veri grubu bir montaj hattındaki bir makinenin

yapacağı hatalarla, montaj hattına olan etkisi olsaydı elde edilecek sonuçlar daha basit ve daha az değişken olurdu.

5.2. Ticari Havayolu Sektörüne Öneriler

Belirli yaş aralığında yüksek kaza frekansı görülüyorsa, o yaş grubu yeniden eğitime altına alınarak kaza yapma riski azaltılabilir. Yahut o yaş grubu mesleğin son yıllarında emekli/iş akdine son verilerek riskli yaşa mensup kişiler sistemden çıkarılabilir.

Belirli bir süre içerisinde aşırı uçuş yapmış veya çok az uçuş yapan pilotlar tespit edilerek bu durum düzeltilmeye çalışılabilir, yönetim çalışanları tarafından yapılacak uygun uçuş planı ile istenmeyen sonuçların oluşması önlenabilir.

Pistin çeşitli durumlarından dolayı meydana gelen kaza ve olayları önlemek için uçuşlar, yönetim çalışanları tarafından farklı tarihlere ve saatlere planlanabilir. İnsan hatalarının; özellikle bazı iniş takımları ve motor tipine bağlı olarak arttığı görüldüğünden, uçak şirketleri yapacakları uçak alımlarında bu bulguları dikkate alarak alımlarına yön verebilirler. Envanterlerinde bulunan uçakların bakımlarının daha hassas yapılması konusunda bakım personeli daha fazla bilgilendirilebilir. Uçak mühendisleri, yeni tasarlayacakları uçaklarda, bu tip uçakların negatif yönlerini göz önünde bulundurabilirler. Işık durumunun insan hatasına etkisini engellemek için, uçuş zamanlarındaki ışık durumu öncelikli değerlendirilebilir.

Öğrenci pilotların özellikle bazı saatler arasında daha çok hata yaptığı tespit edildiğinden, yapılacak eğitim planlamasında bu hususa dikkat edilerek, pilot yetiştirmenin çok maliyetli olduğu havacılık alanının da insan, zaman kaybı ve maddi kayıplar önlenmiş olur.

5.3. Kamu Sektörüne Öneriler

Türkiye'deki ticari uçuş verileri ile oluşturulabilecek bir sistem ile elde edilebilecek Türkiye uçuş koşullarına ait kurallar, internet tabanlı bir program ile havacılık sektörünün ortak kullanımına açılabilir. Devlet destekli bir proje ile anlık uçuş verileri sisteme girilerek, daha önce elde edilen kurallar ile karşılıklı kıyaslanarak; o gün, ay veya yıl içerisindeki uçuşlar da insan hatası meydana gelebilecek olanlar uçuşlar yer ve zaman olarak tespit edilerek ertelenebilir. Böylece birçok kaza ve olay bu bulgular sayesinde önceden öngörülebilir.

Bazı meslek gruplarının özel amaçlı olarak aldıkları lisanslar ile yaptıkları uçuşlarda yaptıkları kazalarda insan hatasının daha fazla gözlendiğinden bu meslek gruplarının

lisans alması veya uçuş yapmaları engellenebilir ya da ortalama eğitim süresi artırılabilir.

Özellikle uçuşun bazı safhalarında daha fazla kaza yapıldığı için, bu safhalarda hava trafik kontrolörlerinin uçuşa daha fazla müdahale etmesi sağlanarak bu uçuş safhaları daha kontrollü hale getirebilir.

Pistlin farklı durumları için meydana gelebilecek kazaları önlemek amacıyla pistler, koşullar değişene değin kapatılarak gelen uçaklar farklı havaalanlarına ya da aynı havaalanının farklı pistlerine yönlendirilebilir.

Uçaklarda kullanılan motorlara sertifika veren ya da kullanım öncesi ölçümlemesini sağlayan bazı bölgelerin, sertifika verme yöntemlerinde yaptıkları hatalar ile insan hatalarına daha çok neden olduklarından, bu bölgelerin daha sıkı denetimlerle ve sahip oldukları uluslar arası standartlar yeniden gözden geçirilerek hata yapması engellenebilir.

Hava Kuvvetleri bünyesinde hali hazırda bulunan veri tabanları, bu çalışmada kullanılan yöntem ile incelenebilir. Elde edilecek bulgular ile Türk Silahlı Kuvvetleri'nin en önemli saldırı gücünün daha etkin ve kayıpsız kullanılması sağlanabilir.

5.4. Gelecek için Çalışmalar

Bu araştırmada uygulanan yöntem, uçak kaza ve olaylarında, kaza etmenlerinin ayrıntılı incelenmesinde bir başlangıç noktasıdır. Bu çalışmanın devamı olarak, kullanılacak algoritmalar başka güvenlik ve kaza verilerine veya birbirinden çok farklı özellikler içeren veri dizilerine uygulanabilir. Ayrıca çalışma, esnasında kullanılan birçok algoritma ve veriler üzerindeki başarımlar değerlendirildiğinden, veri madenciliği konusunda yapılacak çalışmalara da ışık tutacaktır.

Bu çalışmada kullanılan veriler Amerika Birleşik Devletleri'ndeki uçuşları içermektedir. Türkiye'de gerçekleşen sivil veya askeri bütün uçuşlar aynı yöntemle de incelenebilir. Aynı şekilde bu yöntemde de sınıflandırma etmeni olarak, insan hatası değerlendirilmiştir. Değişik özellikler sınıflandırma etmeni olarak kullanılarak, çalışma başka yönlerde incelenebilir ve değerlendirilebilir. Bu çalışma tanımlayıcı (betimsel) yöntemlerden faydalanılarak yapılmıştır, ayrıca stokastik (olasılıklı) tekniklerle değerlendirilebilecek olması gelecekte yeni çalışmaların ortaya çıkmasını sağlayabilir.

Sonuç olarak askeri ve sivil havacılık kazalarının oranlarının azaltılması için harcanacak çaba her zaman yeni bir anlam taşıyacaktır.

KAYNAKLAR

- Aftarczuk, K. 2007. Evaluation of selected data mining algorithms implemented in Medical Decision Support Systems. Master thesis. School of Engineering Blekinge Institute of Technology. 84; 27-32, Sweden.
- Akbulut, S.2006. Veri madenciliği teknikleri ile bir kozmetik markanın ayrılan müşteri analizi ve müşteri segmentasyonu. Yüksek lisans tezi. Gazi üniversitesi, 91; s.4, Ankara.
- Scheffer, T.2001. Finding Association Rules That Trade Support Optimally against Confidence. In: 5th European Conference on Principles of Data Mining and Knowledge Discovery, 424-435
- Anonim. 2002. CAP 715 An introduction to aircraft maintenance engineering human factors for jar 66, Civil Aviation Authority.
- Anonim. 2008. Annual report of the council Doc 9916. International Civil Aviation, Canada.
- Bayam, E., Liebowitz, J., Agresti, W.2005. Older drivers and accidents: A meta analysis and data mining application on traffic accident data, Expert Systems with Applications, Elsevier Ltd., 29; 598–629.
- Bilgi kazancı nedir?
http://ftp.utcluj.ro/pub/users/nedevschi/AV/5_EvaluateClasificatori/entropyEtc.pdf
(03.03.2010)
- Brown, D. Stile, J.R., Gunderson, L., Giras, T.C., 2002. Mining human failure dynamics from accident data using logistic regression and decision trees. IEEE SMC WA2J4
- FAA uçak kaza ve olayları veritabanı
http://av-info.faa.gov/dd_sublevel.asp?Folder=\AID (30.05.2010)
- Frawley, W., Piatetsky-Shapiro, G., Matheus, C. 1992. Knowledge discovery in databases—An overview, AI magazine, AAAI, 29; 57-70.
- Göral, M.A. 2007. Kredi kartı başvuru aşamasında sahtecilik tespiti için bir veri madenciliği modeli.Yüksek lisans tezi, İstanbul Teknik Üniversitesi, 146; s.69, İstanbul
- Graff, A. 2001. Aviation safety, an introduction. Air&Space Europe Volume 3, Issues 3-4, 203-205
- Gürbüz, F., Özbakır, L., Yapıcı, H. 2009a.Classification rule discovery for the aviation incidents resulted in fatality. Knowledge-Based Systems, Elsevier B.V., 22; 622–632
- Gürbüz, F., Özbakır, L., Yapıcı, H.2009b.Türkiye’de bir hava yolu işletmesine ait parça söküm raporlarına ilişkin veri madenciliği uygulaması. Gazi Üniv. Müh. Mim. Fak. Der. Cilt 24, No 1, 73-78
- Han, H.K., Kim, H.S., Sohn, S.Y.2009. Sequential association rules for forecasting failure patterns of aircrafts in Korean Air force, Expert Systems with Applications, Elsevier Ltd., 36; 1129–1133.
- Han, J., Kamber, M. 2000. Data Mining: Concepts and Techniques Chapter 1.Morgan Kaufmann Publishers. 25; 1-26, Canada

- Hand, D., Mannila, H., Smyth, P. 2001. Principles of data mining, The MIT press. 546; 1-21, London, England.
- Harris, Jr., E., Bloedorn, E., Rothleder N.J.1998. Recent experiences with data mining in aviation safety, The MITRE Corporation, 1-7
- İstatistik analizleri el kitabı ve veri madenciliği uygulamaları, 2009 veri madenciliğinde karşılaşılan önemli sorunlar
http://tunedin.entrepreneur.com/tradejournals/article/202562262_5.html (05.05.2010)
- Karakuş, Ö. 2006. Hava aracı kaza ve kırımlarında insan faktörünün araştırılması. Yüksek lisans tezi. Ankara Üniversitesi, 144; s.4, Ankara
- Koyuncuğil A.S., Özgülbaş, N.2009. Veri madenciliği: tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları. Bilişim Teknolojileri Dergisi, CİLT: 2, SAYI: 2, 1-12, Ankara
- Kumar, U., Malik, H., 2003. Analysis of fatal human error aircraft accidents in IAF. Ind J Aerospace Med, 47; 30 – 36.
- Larose, D.2005. Discovering knowledge in data an introduction to data mining. A John Wiley & Sons, Inc., Publication, 222; 1-25, Canada
- Letourneau, S., Famili, F., Matwin, S.1999. Data mining to predict aircraft component replacement. IEEE Intelligent Systems & their applications, 59-66, November/December
- Madani, S. 2009. Mining changes in customer purchasing behaviour – a data mining approach. Master thesis, Lulea University of Technology, 131; s.18, Sweden.
- Marukatat, R. 2006. Structure-based rule selection framework for association rule mining of traffic accident data. IEEE, 781-784, Vol.1, November
- Matyja, D. 2007. Applications of data mining algorithms to analysis of medical data. Master thesis, School of Engineering Blekinge Institute of Technology, 106; 14-22 Sweden.
- Mitra, S., Acharya, T. 2003. Data Mining – Multimedia, Soft Computing and Bioinformatics. A John Wiley & Sons Inc. Publication, 401; 1-30, USA
- Murat, C., 2009, Eurocontrol içinde havacılıkta tek gelişen ülke Türkiye
http://www.aimsa.org.tr/index.php?option=com_content&view=article&id=30%3Aeuro-control-icinde-havacilkta-tek-gelime-goesteren-uelke-tuerkiye&catid=3%3Ahavacilk-haber&lang=tr (25.05.2010)
- Nazeri, Z.2007. Cross-database Analysis to identify relationships between aircraft accidents and incidents, Ph.D. thesis, George Mason University, 160; s.1-8, United States.
- Nazeri. Z., Donohue, G., Sherry, L.2008. Analyzing relationships between aircraft accidents and incidents a data mining approach. International Conference On Research In Air Transportation, Center for Air Transportation Systems Research, 1-5.
- Nazeri, Z., Zhang, J.2002. Mining aviation data to understand impacts of severe weather on airspace system performance. Proceedings of the International Conference on information Technology Coding and Computing, Computer Society IEEE.

NTSB havacılık kaza istatistikleri 1990-2009. [Http://www.nts.gov/aviation/Table2.htm](http://www.nts.gov/aviation/Table2.htm) (15.02.2010)

Ozgan, E.2008.Karayolu araç tipi ve kaza şekli ile kaza sonuçları arasındaki ilişkilerin analizi, Gazi Üniv.Müh.Mim.Fak.Der. Cilt 23, No 1, 97-104.

Öğüdücü, Ş.G., 2008. Veri madenciliği dersi kaynakları
<http://ninova.itu.edu.tr/en/courses/institute-of-informatics/195/bbl-606/ekkaynaklar?g8396> (11.11.2009)

Özkan, Y. 2008.Verit madenciliği yöntemleri.Papatya yayıncılık, 215; İstanbul

Öztekin A., H., 2008, neden hava yolu taşımacılığı, Uçuş Teknisyenleri Derneği aylık havacılık dergisi
<http://www.uted.org/dergi/2008/eylul/bakisacisi/nht.htm> (19.03.2010)

Quinlan, R. 1993. C4.5: Programs for Machine Learning.Publishers, San Mateo, CA.

Pal, R.N., Jain, L. 2005. Advanced techniques in knowledge discovery and data mining with 101 figures. Springer, 252; 1-26, London

Pande, A., Abdel-Aty M. 2007. Market basket analysis of crash data from large jurisdictions and it's potential as a decision support tool. Safety Science, Elsevier Ltd., 47; 145–154

Rapidminer 4.6 Kullanıcı kılavuzu
<http://garr.dl.sourceforge.net/project/yale/1.%20RapidMiner/4.6/rapidminer-4.6-tutorial.pdf> (30.05.2010)

Rapidminer 5.0.6. Veri madenciliği program
<http://rapid-i.com/content/view/167/82/> (10.10.2009)

Rokach, L., Maimon, O. 2008. Data mining with decision trees theory and applications. World Scientific Publishing Co. Pte. Ltd., 242; Singapore

Rygielski, C., Cheng Wang, J., Yen, D.C. 2002. Data mining techniques for customer relationship management.Technology in Society, Elsevier Science Ltd., 24; 483–502

Saraee, M.H., Ehghaghi, Z., Meamarzadeh, H., Zibanezhad, B.2008. Applying data mining in medical data with focus on mortality related to accident in children. IEEE,160-164 12th IEEE International Multitopic Conference, December 23-24

Seifert, J.W. 2006. Data mining and homeland security: an overview. CRS report for congress, Congressional Research Service, The Library of Congress, 1-24.

Shappell,S.,Detwiler,C.,Holcomb,K.,Hackworth,C.,Boquet,A,Wiegmann,M.2006. Human Error and Commercial Aviation Accidents: A Comprehensive, Fine-Grained Analysis Using HFACS, DOT/FAA/AM-06/18, July

Shappell, S.,Wiegmann, D. 2000. The Human Factors Analysis and Classification System–HFACS. Office of Aviation Medicine, U.S. Department of Transportation Federal Aviation Administration, 1-15, Washington

Silahtaroglu, G. 2008. Kavram ve algoritmalarıyla temel veri madenciliği. Papatya yayıncılık eğitim, 172; 83-97, İstanbul.

SPSS 15.0 veri analiz programı

<http://spss-for-windows.software.informer.com/15.0/> (10.12.2009)

Takçı, H., 2008. Veri madenciliğine giriş, G.Y.T.E. veri madenciliği dersi
http://fazliyildirim.com/pdf/364_3.pdf (08.10 .2009)

THY ortaklık verileri yolcu sayısı 2009

http://www.thy.com/tr-TR/corporate/news/press_room/partnership/passanger.aspx
(10.04.2010)

Türk hava kuvvetleri projeleri

<http://www.hvkk.tsk.tr/PageSub/Calismalarimiz/Projeler/F-16Blok50.aspx>
(25.05.2010)

Ulaştırma Bakanı Binali Yıldırım, THY'ye ait uçağın Amsterdam'a inişi esnasında yaşanan üzücü olay ile ilgili basın açıklaması

http://www.ubak.gov.tr/BLSM_WIYS/UBAK/tr/BELGELIK/guncel_haber/20090626_121048_204_1_1541.html (10.04 2010)

Veri madenciliğinin kullanıldığı alanlar, anketler, 2006-2009

<http://www.kdnuggets.com/polls/> 15.02.2010

Wiener, E.1993.Intervention Strategies for the Management of Human Error. NASA Contractor Report 4547, NASA-Ames Research Center, 1-111.

EK-1

Görev

J48 pruned tree

```
c244uçuşplantanımı = NON-UNK
|   c2far = G
|   |   C104.2.UCUSTÜRÜ = N
|   |   |   c102uçuştürü = BU
|   |   |   |   c96uçuşpozisyonu = AL
|   |   |   |   |   c10lokalzaman = O: H (31.0/12.0)
|   |   |   |   |   c10lokalzaman = S: E (26.0/8.0)
|   |   |   |   |   c10lokalzaman = A: H (18.0/6.0)
|   |   |   |   c96uçuşpozisyonu = DS: H (27.0/5.0)
|   |   |   |   c96uçuşpozisyonu = C: H (19.0/5.0)
|   |   |   |   c96uçuşpozisyonu = DP
|   |   |   |   |   c8ayınhaftası = AI: E (1.0)
|   |   |   |   |   c8ayınhaftası = AK: H (5.0/1.0)
|   |   |   |   |   c8ayınhaftası = AC: H (10.0/1.0)
|   |   |   |   |   c8ayınhaftası = AD: E (7.0/1.0)
|   |   |   |   c96uçuşpozisyonu = TO: E (10.0/2.0)
|   |   |   |   c96uçuşpozisyonu = TI: E (3.0/1.0)
|   |   |   c102uçuştürü = PE: E (918.0/317.0)
|   |   |   c102uçuştürü = D: H (1.0)
|   |   |   c102uçuştürü = S
|   |   |   |   c96uçuşpozisyonu = AL: E (157.0/44.0)
|   |   |   |   c96uçuşpozisyonu = DS: H (17.0/6.0)
|   |   |   |   c96uçuşpozisyonu = C: H (50.0/20.0)
|   |   |   |   c96uçuşpozisyonu = DP: E (29.0/8.0)
|   |   |   |   c96uçuşpozisyonu = TO
|   |   |   |   |   c10lokalzaman = O: E (1.0)
|   |   |   |   |   c10lokalzaman = S: H (4.0/1.0)
|   |   |   |   |   c10lokalzaman = A: E (2.0)
|   |   |   |   c96uçuşpozisyonu = TI: E (5.0/2.0)
|   |   |   |   c96uçuşpozisyonu = UKW: H (1.0)
|   |   |   c102uçuştürü = SA: H (3.0)
|   |   |   c102uçuştürü = ATX: E (22.0/6.0)
|   |   |   c102uçuştürü = P: E (1.0)
|   |   |   c102uçuştürü = N: H (1.0)
|   |   |   c102uçuştürü = T: E (1.0)
|   |   |   c102uçuştürü = H: E (3.0/1.0)
|   |   C104.2.UCUSTÜRÜ = B
|   |   |   c96uçuşpozisyonu = AL: E (2813.0/852.0)
|   |   |   c96uçuşpozisyonu = DS
|   |   |   |   c10lokalzaman = O
|   |   |   |   |   c7aylar = K
|   |   |   |   |   |   c6kazayılı = B: E (10.0/3.0)
|   |   |   |   |   |   c6kazayılı = A: H (12.0/6.0)
|   |   |   |   |   |   c6kazayılı = Y: H (5.0/2.0)
|   |   |   |   |   |   c6kazayılı = S: H (14.0/3.0)
|   |   |   |   |   c7aylar = I: E (47.0/17.0)
|   |   |   |   |   c7aylar = Y: E (38.0/13.0)
|   |   |   |   |   c7aylar = S
|   |   |   |   |   |   c6kazayılı = B: H (12.0/4.0)
|   |   |   |   |   |   c6kazayılı = A: E (14.0/4.0)
|   |   |   |   |   |   c6kazayılı = Y: E (14.0/4.0)
```

```

| | | | c10lokalzaman = S
| | | | | c8ayinhaftası = AI: E (25.0/10.0)
| | | | | c8ayinhaftası = AK: H (33.0/15.0)
| | | | | c8ayinhaftası = AC: H (22.0/4.0)
| | | | | c8ayinhaftası = AD
| | | | | | c7aylar = K: H (4.0/2.0)
| | | | | | c7aylar = I: E (10.0/3.0)
| | | | | | c7aylar = Y: H (9.0/2.0)
| | | | | | c7aylar = S: H (6.0/1.0)
| | | | | c10lokalzaman = A: E (66.0/26.0)
| | | c96uçuşpozisyonu = C
| | | | c7aylar = K
| | | | | c10lokalzaman = O
| | | | | | c8ayinhaftası = AI: E (36.0/13.0)
| | | | | | c8ayinhaftası = AK
| | | | | | | c6kazayılı = B: E (9.0/1.0)
| | | | | | | c6kazayılı = A: H (7.0/3.0)
| | | | | | | c6kazayılı = Y: H (8.0/1.0)
| | | | | | | c6kazayılı = S: E (11.0/4.0)
| | | | | | | c8ayinhaftası = AC: H (31.0/14.0)
| | | | | | | c8ayinhaftası = AD: H (28.0/11.0)
| | | | | | c10lokalzaman = S: E (56.0/21.0)
| | | | | | c10lokalzaman = A: E (33.0/11.0)
| | | | | c7aylar = I: E (307.0/137.0)
| | | | | c7aylar = Y
| | | | | | c10lokalzaman = O
| | | | | | | c8ayinhaftası = AI: H (26.0/9.0)
| | | | | | | c8ayinhaftası = AK: H (42.0/17.0)
| | | | | | | c8ayinhaftası = AC: E (38.0/16.0)
| | | | | | | c8ayinhaftası = AD: H (26.0/12.0)
| | | | | | c10lokalzaman = S: H (105.0/49.0)
| | | | | | c10lokalzaman = A: E (58.0/21.0)
| | | | | c7aylar = S: H (178.0/80.0)
| | | c96uçuşpozisyonu = DP: E (465.0/140.0)
| | | c96uçuşpozisyonu = TO: E (170.0/15.0)
| | | c96uçuşpozisyonu = TI
| | | | c6kazayılı = B: H (8.0/1.0)
| | | | c6kazayılı = A: E (2.0)
| | | | c6kazayılı = Y: H (7.0/3.0)
| | | | c6kazayılı = S: E (3.0)
| | | c96uçuşpozisyonu = UKW
| | | | c7aylar = K: E (2.0)
| | | | c7aylar = Y: H (3.0)
| | | | c7aylar = S: H (2.0/1.0)
| | C104.2.UCUSTÜRÜ = S
| | | c102uçuştürü = BU: H (36.0/10.0)
| | | c102uçuştürü = PE: H (2.0)
| | | c102uçuştürü = S: E (3.0/1.0)
| | | c102uçuştürü = ATX: H (5.0/1.0)
| | | c102uçuştürü = T: E (28.0/9.0)
| | C104.2.UCUSTÜRÜ = T
| | | c8ayinhaftası = AI: H (2.0)
| | | c8ayinhaftası = AK: E (1.0)
| | | c8ayinhaftası = AC: H (1.0)
| | | c8ayinhaftası = AD: E (3.0)
| | C104.2.UCUSTÜRÜ = Q: E (49.0/11.0)
| | C104.2.UCUSTÜRÜ = K
| | | c96uçuşpozisyonu = AL: E (18.0/3.0)
| | | c96uçuşpozisyonu = DS: H (8.0/4.0)

```

```

| | | c96uçuşpozisyonu = C: H (11.0/3.0)
| | | c96uçuşpozisyonu = DP: E (4.0/1.0)
| | | c96uçuşpozisyonu = TO: E (2.0)
| | | c96uçuşpozisyonu = TI: H (1.0)
| | C104.2.UCUSTÜRÜ = F
| | | c102uçuştürü = BU: H (2.0/1.0)
| | | c102uçuştürü = PE: E (23.0/8.0)
| | | c102uçuştürü = D: H (1.0)
| | | c102uçuştürü = S
| | | | c10lokalzaman = O: E (22.0/9.0)
| | | | c10lokalzaman = S: E (9.0/2.0)
| | | | c10lokalzaman = A: H (7.0/1.0)
| | | c102uçuştürü = SA: E (1.0)
| | | c102uçuştürü = ATX
| | | | c7aylar = K: H (17.0/5.0)
| | | | c7aylar = I: H (6.0/1.0)
| | | | c7aylar = Y: H (12.0/5.0)
| | | | c7aylar = S: E (8.0/2.0)
| | | c102uçuştürü = P: H (6.0/2.0)
| | C104.2.UCUSTÜRÜ = Y: H (9.0/3.0)
| | C104.2.UCUSTÜRÜ = X: E (1482.0/332.0)
| | C104.2.UCUSTÜRÜ = A: E (19.0/5.0)
| | C104.2.UCUSTÜRÜ = G
| | | c102uçuştürü = BU: H (2.0/1.0)
| | | c102uçuştürü = PE
| | | | c96uçuşpozisyonu = AL: H (24.0/8.0)
| | | | c96uçuşpozisyonu = DS: E (5.0/1.0)
| | | | c96uçuşpozisyonu = C: H (5.0/2.0)
| | | | c96uçuşpozisyonu = DP: E (5.0)
| | | | c96uçuşpozisyonu = TO: E (2.0)
| | | c102uçuştürü = D: H (0.0)
| | | c102uçuştürü = S: H (47.0/10.0)
| | | c102uçuştürü = SA: H (0.0)
| | | c102uçuştürü = ATX: H (4.0/1.0)
| | C104.2.UCUSTÜRÜ = V: E (132.0/35.0)
| | C104.2.UCUSTÜRÜ = CC
| | | c8ayınhaftası = AI
| | | | c102uçuştürü = PE: E (3.0)
| | | | c102uçuştürü = S: H (2.0/1.0)
| | | | c102uçuştürü = ATX: E (0.0)
| | | | c102uçuştürü = T: H (4.0/1.0)
| | | c8ayınhaftası = AK
| | | | c6kazayılı = B: E (8.0/3.0)
| | | | c6kazayılı = A: H (3.0)
| | | | c6kazayılı = Y: H (6.0/1.0)
| | | | c6kazayılı = S: E (1.0)
| | | c8ayınhaftası = AC: E (17.0/3.0)
| | | c8ayınhaftası = AD: H (8.0/2.0)
| | C104.2.UCUSTÜRÜ = M: E (4.0/1.0)
| | C104.2.UCUSTÜRÜ = BB
| | | c8ayınhaftası = AI: H (14.0/4.0)
| | | c8ayınhaftası = AK: E (6.0/2.0)
| | | c8ayınhaftası = AC: E (14.0/4.0)
| | | c8ayınhaftası = AD: E (16.0/6.0)
| | C104.2.UCUSTÜRÜ = H: E (242.0/1.0)
| | C104.2.UCUSTÜRÜ = P
| | | c96uçuşpozisyonu = AL
| | | | c6kazayılı = B: H (2.0)
| | | | c6kazayılı = A: H (1.0)

```

```

| | | | c6kazayılı = Y: E (3.0)
| | | | c6kazayılı = S: H (1.0)
| | | c96uçuşpozisyonu = DS: H (2.0/1.0)
| | | c96uçuşpozisyonu = C: E (16.0/3.0)
| | | c96uçuşpozisyonu = DP: H (1.0)
| | | c96uçuşpozisyonu = TI: H (2.0)
| | | c96uçuşpozisyonu = UKW: E (0.0)
| | C104.2.UCUSTÜRÜ = J: H (17.0/8.0)
| | C104.2.UCUSTÜRÜ = I: E (372.0/4.0)
| | C104.2.UCUSTÜRÜ = AA: E (1.0)
| | C104.2.UCUSTÜRÜ = C: E (1.0)
| c2far = O
| | c102uçuştürü = PE: E (56.0)
| | c102uçuştürü = D: H (19.0/5.0)
| | c102uçuştürü = SA: H (25.0/6.0)
| c2far = T
| | c102uçuştürü = PE: E (2.0)
| | c102uçuştürü = S: H (1.0)
| | c102uçuştürü = SA: H (1.0)
| | c102uçuştürü = ATX
| | | c96uçuşpozisyonu = AL
| | | | c6kazayılı = B: H (77.0/31.0)
| | | | c6kazayılı = A: H (57.0/25.0)
| | | | c6kazayılı = Y
| | | | c7aylar = K: H (17.0/7.0)
| | | | c7aylar = I: E (11.0/3.0)
| | | | c7aylar = Y: H (26.0/10.0)
| | | | c7aylar = S: E (15.0/3.0)
| | | | c6kazayılı = S: H (36.0/12.0)
| | | | c6kazayılı = D: H (0.0)
| | | c96uçuşpozisyonu = DS: H (43.0/8.0)
| | | c96uçuşpozisyonu = C: H (125.0/38.0)
| | | c96uçuşpozisyonu = DP
| | | | c8ayınhaftası = AI: H (11.0/3.0)
| | | | c8ayınhaftası = AK
| | | | c7aylar = K: H (2.0/1.0)
| | | | c7aylar = I: H (4.0)
| | | | c7aylar = Y: H (5.0/1.0)
| | | | c7aylar = S: E (6.0/1.0)
| | | | c8ayınhaftası = AC: E (14.0/6.0)
| | | | c8ayınhaftası = AD
| | | | c10lokalzaman = O: H (8.0/2.0)
| | | | c10lokalzaman = S: H (2.0)
| | | | c10lokalzaman = A: E (7.0)
| | | c96uçuşpozisyonu = TO: E (25.0/8.0)
| | | c96uçuşpozisyonu = TI: H (13.0/4.0)
| | | c96uçuşpozisyonu = UKW: H (1.0)
| | c102uçuştürü = P: H (1.0)
| | c102uçuştürü = N: E (1.0)
| | c102uçuştürü = T: E (33.0/1.0)
| c2far = P
| | C104.2.UCUSTÜRÜ = K: H (70.0/3.0)
| | C104.2.UCUSTÜRÜ = J: E (6.0)
| c2far = A: E (347.0/140.0)
| c2far = U
| | c10lokalzaman = O: H (3.0)
| | c10lokalzaman = S: E (7.0/2.0)
| | c10lokalzaman = A: E (2.0)
| c2far = H

```

```

| | c102uçuştürü = PE
| | | c8ayınhaftası = AI
| | | | c96uçuşpozisyonu = AL: H (2.0)
| | | | c96uçuşpozisyonu = C: E (4.0/1.0)
| | | | c96uçuşpozisyonu = UKW: H (1.0)
| | | c8ayınhaftası = AK: H (6.0)
| | | c8ayınhaftası = AC
| | | | c10lokalzaman = O: H (6.0/1.0)
| | | | c10lokalzaman = S: E (8.0/1.0)
| | | | c10lokalzaman = A: E (1.0)
| | | c8ayınhaftası = AD: H (5.0/2.0)
| | c102uçuştürü = S: E (6.0)
| | c102uçuştürü = SA: H (0.0)
| | c102uçuştürü = ATX: H (0.0)
| | c102uçuştürü = P: H (1.0)
| c2far = F: E (17.0/1.0)
| c2far = C: E (3.0/1.0)
c244uçuşplantanımı = IFR
| c2far = G
| | c96uçuşpozisyonu = AL: E (103.0/26.0)
| | c96uçuşpozisyonu = DS
| | | c6kazayılı = B: H (8.0/2.0)
| | | c6kazayılı = A
| | | | c10lokalzaman = O: E (5.0/1.0)
| | | | c10lokalzaman = S: H (2.0)
| | | | c10lokalzaman = A: H (3.0/1.0)
| | | c6kazayılı = Y: E (9.0/2.0)
| | | c6kazayılı = S: E (3.0/1.0)
| | c96uçuşpozisyonu = C: H (81.0/32.0)
| | c96uçuşpozisyonu = DP: E (21.0/3.0)
| | c96uçuşpozisyonu = TO: E (5.0/1.0)
| c2far = O: H (656.0/56.0)
| c2far = T: E (32.0/9.0)
| c2far = C: H (2.0/1.0)
c244uçuşplantanımı = VFR: E (88.0/25.0)
c244uçuşplantanımı = VFF
| c10lokalzaman = O: H (7.0/2.0)
| c10lokalzaman = S: E (4.0)
| c10lokalzaman = A: H (4.0/1.0)
c244uçuşplantanımı = DFR: H (1.0)
c244uçuşplantanımı = CVF: E (2.0)

```

İnsan

J48 pruned tree

c47diğerrating = D

```

| C233COMDPLTölüm = N
| | C52UZMANLIKİŞ = O: E (3479.0/1273.0)
| | C52UZMANLIKİŞ = L: H (1.0)
| | C52UZMANLIKİŞ = D: E (2.0)
| | C52UZMANLIKİŞ = G: E (3.0/1.0)
| | C52UZMANLIKİŞ = F: H (2.0)
| | C52UZMANLIKİŞ = K
| | | c55son90gün = B

```

```

| | | | c50pltyaş = E: H (196.0/91.0)
| | | | c50pltyaş = Y: E (627.0/278.0)
| | | | c50pltyaş = O: H (120.0/51.0)
| | | | c50pltyaş = G: H (69.0/28.0)
| | | c55son90gün = N: H (574.0/218.0)
| | | c55son90gün = E: E (63.0/26.0)
| | | c55son90gün = X: E (2.0)
| | | c55son90gün = XX: H (1.0)

```

```

| C233COMDPLTölüm = Y: E (434.0/71.0)
c47diğerrating = A: E (3157.0/904.0)
c47diğerrating = X: E (6.0/2.0)
c47diğerrating = G: E (377.0/126.0)
c47diğerrating = C: E (2308.0/49.0)
c47diğerrating = B: E (623.0/54.0)
c47diğerrating = E: H (64.0/31.0)
c47diğerrating = F: H (40.0/16.0)
c47diğerrating = H
| c50pltyaş = E: H (2.0)
| c50pltyaş = Y: H (8.0/3.0)
| c50pltyaş = O: E (1.0)
| c50pltyaş = G: E (2.0)

```

Kaza

J48 pruned tree

```

C82HANGİİNSANVAR = A
| C184DESTEKLEYİCİOPERASYONELKOD = A: E (3761.0/42.0)
| C184DESTEKLEYİCİOPERASYONELKOD = F
| | c61toplaminsan = A: E (2362.0/5.0)
| | c61toplaminsan = N
| | | C80YARDIMCINEDENFAKTÖRÜ = V: E (5.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = Z: E (1.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = B: E (1.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = S: H (22.0/2.0)
| | c61toplaminsan = Y: E (12.0)
| C184DESTEKLEYİCİOPERASYONELKOD = E: E (2896.0/5.0)
| C184DESTEKLEYİCİOPERASYONELKOD = I
| | C80YARDIMCINEDENFAKTÖRÜ = V: E (227.0)
| | C80YARDIMCINEDENFAKTÖRÜ = Z: E (23.0)
| | C80YARDIMCINEDENFAKTÖRÜ = A: E (1.0)
| | C80YARDIMCINEDENFAKTÖRÜ = O: E (2.0)
| | C80YARDIMCINEDENFAKTÖRÜ = B: E (1.0)
| | C80YARDIMCINEDENFAKTÖRÜ = E: E (68.0/2.0)
| | C80YARDIMCINEDENFAKTÖRÜ = T
| | | c98hasarkod = S: E (2.0)
| | | c98hasarkod = M: H (6.0)
| | | c98hasarkod = D: E (1.0)
| | | c98hasarkod = N: H (1.0)
| | C80YARDIMCINEDENFAKTÖRÜ = S: H (4.0)
| | C80YARDIMCINEDENFAKTÖRÜ = Q: H (24.0/1.0)
| | C80YARDIMCINEDENFAKTÖRÜ = R: H (2.0)
| | C80YARDIMCINEDENFAKTÖRÜ = G: E (1.0)
| | C80YARDIMCINEDENFAKTÖRÜ = H: E (2.0)
| | C80YARDIMCINEDENFAKTÖRÜ = Y: E (3.0)

```

```

| | C80YARDIMCINEDENFAKTÖRÜ = K: E (0.0)
| | C80YARDIMCINEDENFAKTÖRÜ = D: E (1.0)
| | C80YARDIMCINEDENFAKTÖRÜ = C: E (1.0)
| C184DESTEKLEYİCİOPERASYONELKOD = T: E (1.0)
| C184DESTEKLEYİCİOPERASYONELKOD = S
| | c78birincinedenfaktörü = MO: E (17.0)
| | c78birincinedenfaktörü = HO: E (73.0/20.0)
| | c78birincinedenfaktörü = OT: H (1.0)
| | c78birincinedenfaktörü = II: E (8.0)
| | c78birincinedenfaktörü = GN: E (7.0/2.0)
| | c78birincinedenfaktörü = OS: E (9.0)
| | c78birincinedenfaktörü = GC: E (138.0/9.0)
| | c78birincinedenfaktörü = AS: E (19.0/2.0)
| | c78birincinedenfaktörü = PF: E (9.0/2.0)
| | c78birincinedenfaktörü = WX: E (7.0)
| | c78birincinedenfaktörü = IO: E (80.0/5.0)
| | c78birincinedenfaktörü = S: H (1.0)
| | c78birincinedenfaktörü = FT: E (12.0)
| | c78birincinedenfaktörü = PS: E (30.0)
| | c78birincinedenfaktörü = WF: E (1.0)
| | c78birincinedenfaktörü = GI: E (3.0)
| | c78birincinedenfaktörü = LC
| | | c98hasarkod = S: H (7.0)
| | | c98hasarkod = M: H (2.0)
| | | c98hasarkod = D: E (11.0/3.0)
| | | c98hasarkod = N: H (2.0)
| | c78birincinedenfaktörü = IA: E (3.0)
| | c78birincinedenfaktörü = PI: H (1.0)
| | c78birincinedenfaktörü = FA: H (2.0)
| | c78birincinedenfaktörü = TC: E (22.0/3.0)
| | c78birincinedenfaktörü = SN
| | | C80YARDIMCINEDENFAKTÖRÜ = V: H (5.0/2.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = B: E (4.0/1.0)
| | c78birincinedenfaktörü = ST: E (2.0)
| | c78birincinedenfaktörü = LO: E (41.0/12.0)
| | c78birincinedenfaktörü = IT: E (1.0)
| | c78birincinedenfaktörü = CH: E (9.0)
| | c78birincinedenfaktörü = PW: E (3.0/1.0)
| | c78birincinedenfaktörü = US: E (9.0)
| | c78birincinedenfaktörü = IP: E (2.0)
| | c78birincinedenfaktörü = AU: E (2.0)
| | c78birincinedenfaktörü = GE: E (5.0)
| | c78birincinedenfaktörü = HA: E (12.0/2.0)
| | c78birincinedenfaktörü = SD: H (2.0)
| | c78birincinedenfaktörü = FP: E (10.0/1.0)
| | c78birincinedenfaktörü = UN: E (3.0/1.0)
| | c78birincinedenfaktörü = GF: E (2.0)
| | c78birincinedenfaktörü = OM
| | | C80YARDIMCINEDENFAKTÖRÜ = V: E (55.0/23.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = Z: H (7.0/1.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = A: H (3.0/1.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = B: E (2.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = I: H (3.0/1.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = E: H (7.0/2.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = T: E (1.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = G: E (2.0/1.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = H: H (1.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = Y: E (1.0)
| | | C80YARDIMCINEDENFAKTÖRÜ = D: H (1.0)

```

| | | C80YARDIMCINEDENFAKTÖRÜ = C: H (1.0)
 | | | C80YARDIMCINEDENFAKTÖRÜ = P: H (1.0)
 | | c78birincinedenfaktörü = PC
 | | | c94kazakodutürü = AC: H (3.0/1.0)
 | | | c94kazakodutürü = IA: H (2.0)
 | | | c94kazakodutürü = IB: E (4.0)
 | | | c94kazakodutürü = SE: H (3.0/1.0)
 | | | c94kazakodutürü = KX: H (1.0)
 | | c78birincinedenfaktörü = BW: H (37.0/1.0)
 | | c78birincinedenfaktörü = WS: H (20.0/4.0)
 | | c78birincinedenfaktörü = UD: H (9.0)
 | | c78birincinedenfaktörü = PB: E (3.0)
 | | c78birincinedenfaktörü = UT: E (32.0/14.0)
 | | c78birincinedenfaktörü = FX: E (15.0/4.0)
 | | c78birincinedenfaktörü = SP: E (1.0)
 | | c78birincinedenfaktörü = IS: E (6.0/3.0)
 | | c78birincinedenfaktörü = PO: E (3.0)
 | | c78birincinedenfaktörü = ES: H (1.0)
 | | c78birincinedenfaktörü = PP: E (1.0)
 | | c78birincinedenfaktörü = SR: H (1.0)
 | | c78birincinedenfaktörü = LR: E (2.0)
 | | c78birincinedenfaktörü = FO: E (3.0/1.0)
 | | c78birincinedenfaktörü = PL: E (1.0)
 | | c78birincinedenfaktörü = DW: E (7.0/3.0)
 | | c78birincinedenfaktörü = DC: E (6.0)
 | | c78birincinedenfaktörü = XL: E (3.0)
 | | c78birincinedenfaktörü = RI: E (4.0)
 | | c78birincinedenfaktörü = HG
 | | | C80YARDIMCINEDENFAKTÖRÜ = V: E (3.0/1.0)
 | | | C80YARDIMCINEDENFAKTÖRÜ = T: H (2.0)
 | | | C80YARDIMCINEDENFAKTÖRÜ = G: E (1.0)
 | | c78birincinedenfaktörü = OC: E (2.0)
 | | c78birincinedenfaktörü = UA: E (4.0/1.0)
 | | c78birincinedenfaktörü = GL: E (1.0)
 | | c78birincinedenfaktörü = MA: E (11.0/2.0)
 | | c78birincinedenfaktörü = FR: E (2.0)
 | | c78birincinedenfaktörü = RS: E (3.0)
 | | c78birincinedenfaktörü = MM: E (1.0)
 | | c78birincinedenfaktörü = IE: E (1.0)
 | | c78birincinedenfaktörü = PA: E (1.0)
 | | c78birincinedenfaktörü = WO: H (1.0)
 | C184DESTEKLEYİCİOPERASYONELKOD = J: H (343.0/22.0)
 | C184DESTEKLEYİCİOPERASYONELKOD = B: E (76.0/4.0)
 | C184DESTEKLEYİCİOPERASYONELKOD = G
 | | C80YARDIMCINEDENFAKTÖRÜ = V: E (59.0/1.0)
 | | C80YARDIMCINEDENFAKTÖRÜ = Z: E (10.0)
 | | C80YARDIMCINEDENFAKTÖRÜ = I: E (1.0)
 | | C80YARDIMCINEDENFAKTÖRÜ = E: E (1.0)
 | | C80YARDIMCINEDENFAKTÖRÜ = T: H (3.0)
 | | C80YARDIMCINEDENFAKTÖRÜ = S: H (3.0)
 | | C80YARDIMCINEDENFAKTÖRÜ = W: E (2.0)
 | | C80YARDIMCINEDENFAKTÖRÜ = Y: E (1.0)
 | C184DESTEKLEYİCİOPERASYONELKOD = Q: E (18.0)
 | C184DESTEKLEYİCİOPERASYONELKOD = V: E (2.0/1.0)
 | C184DESTEKLEYİCİOPERASYONELKOD = U
 | | C80YARDIMCINEDENFAKTÖRÜ = V: E (10.0/5.0)
 | | C80YARDIMCINEDENFAKTÖRÜ = Z: E (2.0)
 | | C80YARDIMCINEDENFAKTÖRÜ = E: E (1.0)
 | | C80YARDIMCINEDENFAKTÖRÜ = R: H (20.0/1.0)

```

| | C80YARDIMCINEDENFAKTÖRÜ = G: E (1.0)
| | C80YARDIMCINEDENFAKTÖRÜ = H: H (1.0)
| | C80YARDIMCINEDENFAKTÖRÜ = F: E (2.0)
| C184DESTEKLEYİCİOPERASYONELKOD = M: E (6.0)
| C184DESTEKLEYİCİOPERASYONELKOD = C
| | c78birincinedenfaktörü = HO: E (3.0)
| | c78birincinedenfaktörü = OS: E (2.0)
| | c78birincinedenfaktörü = GC: E (2.0)
| | c78birincinedenfaktörü = AS: E (2.0)
| | c78birincinedenfaktörü = PF: E (1.0)
| | c78birincinedenfaktörü = IO: E (11.0)
| | c78birincinedenfaktörü = LC: H (5.0)
| | c78birincinedenfaktörü = TC: E (1.0)
| | c78birincinedenfaktörü = AU: H (1.0)
| | c78birincinedenfaktörü = OM: H (4.0)
| | c78birincinedenfaktörü = PC: E (1.0)
| | c78birincinedenfaktörü = UT: H (1.0)
| | c78birincinedenfaktörü = IS: E (1.0)
| | c78birincinedenfaktörü = LA
| | | toplamölen = Y: H (3.0/1.0)
| | | toplamölen = A: E (2.0)
| | c78birincinedenfaktörü = PO: E (1.0)
| | c78birincinedenfaktörü = DW: E (2.0)
| | c78birincinedenfaktörü = XL: E (1.0)
| | c78birincinedenfaktörü = HG: E (1.0)
| | c78birincinedenfaktörü = ME: H (4.0)
| | c78birincinedenfaktörü = MA: H (2.0)
| C184DESTEKLEYİCİOPERASYONELKOD = H: E (135.0/8.0)
| C184DESTEKLEYİCİOPERASYONELKOD = W: E (6.0)
| C184DESTEKLEYİCİOPERASYONELKOD = L: H (18.0/5.0)
| C184DESTEKLEYİCİOPERASYONELKOD = N: E (4.0)
| C184DESTEKLEYİCİOPERASYONELKOD = P: E (2.0/1.0)
| C184DESTEKLEYİCİOPERASYONELKOD = D: E (7.0/1.0)
| C184DESTEKLEYİCİOPERASYONELKOD = K: E (3.0/1.0)
C82HANGİİNSANVAR = M: H (144.0/2.0)
C82HANGİİNSANVAR = P: H (198.0/14.0)
C82HANGİİNSANVAR = H: H (63.0/3.0)
C82HANGİİNSANVAR = L: H (24.0/1.0)
C82HANGİİNSANVAR = Q: H (11.0/2.0)
C82HANGİİNSANVAR = E
| toplamölen = Y: H (37.0/1.0)
| toplamölen = A: E (2.0)
C82HANGİİNSANVAR = G
| c61toplaminsan = A: E (5.0)
| c61toplaminsan = N: H (28.0)
| c61toplaminsan = Ç: H (6.0)
C82HANGİİNSANVAR = O: H (7.0)

```

Makine

J48 pruned tree

```

c145ağırlık = OV
| c157iniştakım = RT
| | C155MOTORKOD = T: H (313.0/135.0)
| | C155MOTORKOD = O: H (7.0/3.0)
| | C155MOTORKOD = R: H (707.0/263.0)
| | C155MOTORKOD = U

```

```

|   |   |   C27BÖLGE SERTİFİKA = EU: E (2.0)
|   |   |   C27BÖLGE SERTİFİKA = NM: E (38.0/14.0)
|   |   |   C27BÖLGE SERTİFİKA = WP: H (19.0/8.0)
|   |   C155MOTORKOD = V
|   |   |   C27BÖLGE SERTİFİKA = NM: E (5.0/2.0)
|   |   |   C27BÖLGE SERTİFİKA = SW: E (1.0)
|   |   |   C27BÖLGE SERTİFİKA = WP: H (8.0/1.0)
|   |   C155MOTORKOD = J: E (10.0/3.0)
|   |   C155MOTORKOD = W: H (1.0)
|   c157iniştakım = RC: E (27.0/9.0)
|   c157iniştakım = NT: H (1.0)
c145ağırlık = UN: E (13666.0/3400.0)

```

Ortam 1

J48 pruned tree

```

c110ışıkdurumu = D
|   c75kazadakaçuçakvar = N: E (14511.0/3970.0)
|   c75kazadakaçuçakvar = T: H (2.0/1.0)
|   c75kazadakaçuçakvar = F: E (43.0/21.0)
|   c75kazadakaçuçakvar = S: H (43.0/12.0)
c110ışıkdurumu = N
|   c117pistdurumu = DRY: H (591.0/256.0)
|   c117pistdurumu = WET: E (40.0/17.0)
|   c117pistdurumu = SNP: E (7.0/1.0)
|   c117pistdurumu = ICE: E (381.0/108.0)
|   c117pistdurumu = WTR: H (2.0)
|   c117pistdurumu = ICP: E (0.0)
|   c117pistdurumu = SNU: E (12.0/5.0)
c110ışıkdurumu = DU: E (28.0/10.0)
c110ışıkdurumu = DA: H (10.0/5.0)

```

Ortam 2

J48 pruned tree

```

C1082.UÇUŞDURUM = O: E (60.0/26.0)
C1082.UÇUŞDURUM = V: E (480.0/144.0)
C1082.UÇUŞDURUM = G: E (1367.0/235.0)
C1082.UÇUŞDURUM = H: E (133.0/51.0)
C1082.UÇUŞDURUM = U: H (80.0/26.0)
C1082.UÇUŞDURUM = Q: E (4398.0/21.0)
C1082.UÇUŞDURUM = F: E (109.0/14.0)
C1082.UÇUŞDURUM = T: E (141.0/57.0)
C1082.UÇUŞDURUM = I: E (9.0/1.0)
C1082.UÇUŞDURUM = N: E (51.0/19.0)
C1082.UÇUŞDURUM = M: E (16.0/6.0)
C1082.UÇUŞDURUM = L: E (40.0/1.0)
C1082.UÇUŞDURUM = C: E (3.0)
C1082.UÇUŞDURUM = B: H (1.0)
C1082.UÇUŞDURUM = S
|   C106UÇUŞDURUM = C: H (1111.0/447.0)
|   C106UÇUŞDURUM = B: E (7573.0/3119.0)
|   C106UÇUŞDURUM = A: H (91.0/26.0)
C1082.UÇUŞDURUM = J

```

```
| C106UÇUŞDURUM = C: H (4.0/1.0)
| C106UÇUŞDURUM = B: H (0.0)
| C106UÇUŞDURUM = A: E (3.0/1.0)
```

EK-2

Görev

Association Rules

[c244uçuşplantanımı = NON-UNK, c96uçuşpozisyonu = AL, c10lokalzaman = O] --> [c2far = G] (confidence: 0.950)
[c244uçuşplantanımı = NON-UNK, c100insanhatası = E, c96uçuşpozisyonu = AL] --> [c2far = G] (confidence: 0.951)
[c102uçuştürü = PE, c10lokalzaman = O] --> [c244uçuşplantanımı = NON-UNK, c2far = G] (confidence: 0.954)
[c96uçuşpozisyonu = AL, c10lokalzaman = O] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.956)
[c100insanhatası = E] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.957)
[c2far = G, c100insanhatası = H] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.959)
[c100insanhatası = E, c96uçuşpozisyonu = AL] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.962)
[c2far = G, c102uçuştürü = PE] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.963)
[c102uçuştürü = PE] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.963)
[c102uçuştürü = PE, c96uçuşpozisyonu = AL] --> [c244uçuşplantanımı = NON-UNK, c2far = G] (confidence: 0.963)
[c2far = G, c100insanhatası = E, c102uçuştürü = PE] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.964)
[c100insanhatası = E, c102uçuştürü = PE] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.964)
[c2far = G] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.964)
[c100insanhatası = E, c10lokalzaman = O] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.964)
[c2far = G, c6kazayılı = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.966)
[c2far = G, c102uçuştürü = PE, c10lokalzaman = O] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.966)
[c102uçuştürü = PE, c10lokalzaman = O] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.966)
[c2far = G, c10lokalzaman = O] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.967)
[c2far = G, c100insanhatası = E] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.967)
[C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c2far = G, c102uçuştürü = PE] (confidence: 0.967)
[c2far = G, c10lokalzaman = S] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.967)
[C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c2far = G] (confidence: 0.968)
[c102uçuştürü = PE, c96uçuşpozisyonu = AL] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.968)
[c102uçuştürü = PE, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c2far = G] (confidence: 0.968)
[c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c2far = G, c102uçuştürü = PE] (confidence: 0.968)
[c2far = G, c102uçuştürü = PE, c96uçuşpozisyonu = AL] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.968)
[c2far = G, c7aylar = I] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.969)
[C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c102uçuştürü

= PE] (confidence: 0.969)
 [c2far = G, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE] (confidence: 0.969)
 [c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c2far = G] (confidence: 0.969)
 [c100insanhatası = E, c102uçuştürü = PE, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c2far = G] (confidence: 0.969)
 [c2far = G, c102uçuştürü = PE, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.970)
 [c2far = G, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.970)
 [c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c2far = G, c102uçuştürü = PE] (confidence: 0.970)
 [c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE] (confidence: 0.970)
 [c102uçuştürü = PE, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.970)
 [C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.970)
 [c2far = G, c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE] (confidence: 0.970)
 [c102uçuştürü = PE, c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c2far = G] (confidence: 0.970)
 [c2far = G, c100insanhatası = E, c10lokalzaman = O] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.970)
 [c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c2far = G] (confidence: 0.970)
 [c2far = G, c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE] (confidence: 0.970)
 [c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE] (confidence: 0.970)
 [c2far = G, c102uçuştürü = PE, c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.971)
 [c2far = G, c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.971)
 [c102uçuştürü = PE, c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.971)
 [c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.971)
 [c2far = G, c100insanhatası = E, c102uçuştürü = PE, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.971)
 [c2far = G, c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.971)
 [c100insanhatası = E, c102uçuştürü = PE, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.971)
 [c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.971)
 [c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c2far = G] (confidence: 0.972)
 [c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c2far = G, c102uçuştürü = PE] (confidence: 0.972)
 [c102uçuştürü = PE, c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c2far = G] (confidence: 0.972)
 [c2far = G, c100insanhatası = E, c96uçuşpozisyonu = AL] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.973)
 [c2far = G, c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.973)
 [c2far = G, c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE] (confidence: 0.973)

[c2far = G, c102uçuştürü = PE, c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.973)
 [c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.973)
 [c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE] (confidence: 0.973)
 [c102uçuştürü = PE, c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.973)
 [c2far = G, c96uçuşpozisyonu = AL] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.973)
 [c2far = G, c96uçuşpozisyonu = AL, c10lokalzaman = O] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.976)
 [c2far = G, c7aylar = Y] --> [c244uçuşplantanımı = NON-UNK] (confidence: 0.976)
 [c244uçuşplantanımı = NON-UNK, c100insanhatası = E, c102uçuştürü = PE] --> [c2far = G] (confidence: 0.980)
 [c100insanhatası = E, c102uçuştürü = PE] --> [c2far = G] (confidence: 0.980)
 [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE] --> [c2far = G] (confidence: 0.983)
 [c102uçuştürü = PE] --> [c2far = G] (confidence: 0.983)
 [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE, c10lokalzaman = O] --> [c2far = G] (confidence: 0.987)
 [c102uçuştürü = PE, c10lokalzaman = O] --> [c2far = G] (confidence: 0.987)
 [c102uçuştürü = PE, c96uçuşpozisyonu = AL] --> [c2far = G] (confidence: 0.995)
 [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE, c96uçuşpozisyonu = AL] --> [c2far = G] (confidence: 0.995)
 [c100insanhatası = E, c102uçuştürü = PE, c96uçuşpozisyonu = AL] --> [c2far = G] (confidence: 0.996)
 [c244uçuşplantanımı = NON-UNK, c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c2far = G, c102uçuştürü = PE] (confidence: 0.997)
 [c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c2far = G, c102uçuştürü = PE] (confidence: 0.997)
 [c244uçuşplantanımı = NON-UNK, C104.2.UCUSTÜRÜ = B] --> [c2far = G, c102uçuştürü = PE] (confidence: 0.998)
 [C104.2.UCUSTÜRÜ = B] --> [c2far = G, c102uçuştürü = PE] (confidence: 0.998)
 [c244uçuşplantanımı = NON-UNK, c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.998)
 [c244uçuşplantanımı = NON-UNK, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.998)
 [c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.998)
 [C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.998)
 [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.998)
 [c102uçuştürü = PE, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.998)
 [c244uçuşplantanımı = NON-UNK, c100insanhatası = E, c102uçuştürü = PE, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.998)
 [c100insanhatası = E, c102uçuştürü = PE, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.998)
 [c244uçuşplantanımı = NON-UNK, c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 0.999)
 [c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 0.999)
 [c244uçuşplantanımı = NON-UNK, c10lokalzaman = O, C104.2.UCUSTÜRÜ = B]

--> [c2far = G, c102uçuştürü = PE] (confidence: 0.999)
 [c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c2far = G, c102uçuştürü = PE] (confidence: 0.999)
 [c244uçuşplantanımı = NON-UNK, c2far = G, c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 0.999)
 [c2far = G, c100insanhatası = E, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 0.999)
 [c244uçuşplantanımı = NON-UNK, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 0.999)
 [C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 0.999)
 [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE, c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.999)
 [c244uçuşplantanımı = NON-UNK, c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.999)
 [c102uçuştürü = PE, c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.999)
 [c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.999)
 [c244uçuşplantanımı = NON-UNK, c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.999)
 [c244uçuşplantanımı = NON-UNK, c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c2far = G, c102uçuştürü = PE] (confidence: 0.999)
 [c244uçuşplantanımı = NON-UNK, c102uçuştürü = PE, c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.999)
 [c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.999)
 [c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c2far = G, c102uçuştürü = PE] (confidence: 0.999)
 [c102uçuştürü = PE, c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c2far = G] (confidence: 0.999)
 [c244uçuşplantanımı = NON-UNK, c2far = G, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 0.999)
 [c2far = G, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 0.999)
 [c244uçuşplantanımı = NON-UNK, c2far = G, c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 1.000)
 [c244uçuşplantanımı = NON-UNK, c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 1.000)
 [c2far = G, c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 1.000)
 [c10lokalzaman = O, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 1.000)
 [c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 1.000)
 [c244uçuşplantanımı = NON-UNK, c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 1.000)
 [c2far = G, c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 1.000)
 [c244uçuşplantanımı = NON-UNK, c2far = G, c96uçuşpozisyonu = AL, C104.2.UCUSTÜRÜ = B] --> [c102uçuştürü = PE] (confidence: 1.000)

İnsan

Association Rules

[C49KALİTE = S, C52UZMANLIKİŞ = O, c50pltyaş = Y] --> [c55son90gün = B] (confidence: 0.952)
 [C49KALİTE = S, C233COMDPLTölüm = N, C52UZMANLIKİŞ = O, c50pltyaş = Y] --> [c55son90gün = B] (confidence: 0.953)

[C233COMDPLTölüm = N, c55son90gün = B, C52UZMANLIKİŞ = O, c50pltyaş = Y] --> [C49KALİTE = S] (confidence: 0.954)
 [c55son90gün = B, c47diğerrating = D] --> [C49KALİTE = S] (confidence: 0.954)
 [C233COMDPLTölüm = N, C52UZMANLIKİŞ = O, c50pltyaş = Y] --> [c55son90gün = B] (confidence: 0.955)
 [C233COMDPLTölüm = N, C52UZMANLIKİŞ = O, c50pltyaş = Y] --> [C49KALİTE = S] (confidence: 0.955)
 [c53uçaktoplamaçuş = B, c50pltyaş = Y] --> [c55son90gün = B] (confidence: 0.957)
 [C52UZMANLIKİŞ = O, c50pltyaş = Y] --> [c55son90gün = B] (confidence: 0.957)
 [C233COMDPLTölüm = N, c55son90gün = B, c50pltyaş = Y] --> [C49KALİTE = S] (confidence: 0.957)
 [C49KALİTE = S, c50pltyaş = Y] --> [C233COMDPLTölüm = N] (confidence: 0.960)
 [C233COMDPLTölüm = N, c50pltyaş = Y] --> [C49KALİTE = S] (confidence: 0.960)
 [C49KALİTE = S, c55son90gün = B, c50pltyaş = Y] --> [C233COMDPLTölüm = N] (confidence: 0.961)
 [c47diğerrating = D] --> [C49KALİTE = S] (confidence: 0.962)
 [C233COMDPLTölüm = N, c55son90gün = B, C52UZMANLIKİŞ = O] --> [C49KALİTE = S] (confidence: 0.963)
 [C233COMDPLTölüm = N, c55son90gün = B, C52UZMANLIKİŞ = O, c100insanhataşı = E] --> [C49KALİTE = S] (confidence: 0.963)
 [C233COMDPLTölüm = N, C41CPMDPLTSERTİFİKA = FA] --> [C49KALİTE = S, C52UZMANLIKİŞ = O] (confidence: 0.963)
 [C49KALİTE = S, C52UZMANLIKİŞ = O, c50pltyaş = Y] --> [C233COMDPLTölüm = N] (confidence: 0.964)
 [C233COMDPLTölüm = N, c55son90gün = B, C41CPMDPLTSERTİFİKA = FA] --> [C49KALİTE = S, C52UZMANLIKİŞ = O] (confidence: 0.964)
 [C233COMDPLTölüm = N, c55son90gün = B, C52UZMANLIKİŞ = O, c53uçaktoplamaçuş = B] --> [C49KALİTE = S] (confidence: 0.964)
 [C233COMDPLTölüm = N, c55son90gün = B, C52UZMANLIKİŞ = O, c100insanhataşı = E, c53uçaktoplamaçuş = B] --> [C49KALİTE = S] (confidence: 0.965)
 [C49KALİTE = S, c55son90gün = B, C52UZMANLIKİŞ = O, c50pltyaş = Y] --> [C233COMDPLTölüm = N] (confidence: 0.965)
 [C233COMDPLTölüm = N, C52UZMANLIKİŞ = O, C41CPMDPLTSERTİFİKA = FA] --> [C49KALİTE = S] (confidence: 0.965)
 [C233COMDPLTölüm = N, C41CPMDPLTSERTİFİKA = FA] --> [C49KALİTE = S] (confidence: 0.965)
 [C233COMDPLTölüm = N, c55son90gün = B, C52UZMANLIKİŞ = O, C41CPMDPLTSERTİFİKA = FA] --> [C49KALİTE = S] (confidence: 0.965)
 [C233COMDPLTölüm = N, c55son90gün = B, C41CPMDPLTSERTİFİKA = FA] --> [C49KALİTE = S] (confidence: 0.965)
 [C233COMDPLTölüm = N, c55son90gün = B, c100insanhataşı = E] --> [C49KALİTE = S] (confidence: 0.966)
 [C233COMDPLTölüm = N, C52UZMANLIKİŞ = O, c100insanhataşı = E] --> [C49KALİTE = S] (confidence: 0.966)
 [C233COMDPLTölüm = N, C52UZMANLIKİŞ = O] --> [C49KALİTE = S] (confidence: 0.966)
 [C233COMDPLTölüm = N, c55son90gün = B, c100insanhataşı = E, c53uçaktoplamaçuş = B] --> [C49KALİTE = S] (confidence: 0.966)
 [C233COMDPLTölüm = N, c55son90gün = B] --> [C49KALİTE = S] (confidence: 0.966)
 [C233COMDPLTölüm = N, C52UZMANLIKİŞ = O, c53uçaktoplamaçuş = B] --> [C49KALİTE = S] (confidence: 0.966)
 [C233COMDPLTölüm = N, c55son90gün = B, c53uçaktoplamaçuş = B] --> [C49KALİTE = S] (confidence: 0.966)

[C49KALİTE = S] (confidence: 0.966)
 [C233COMDPLTölüm = N, C52UZMANLIKİŞ = O, c100insanhatası = E, c53uçaktoplamaçuş = B] --> [C49KALİTE = S] (confidence: 0.967)
 [C233COMDPLTölüm = N, c100insanhatası = E, c53uçaktoplamaçuş = B] --> [C49KALİTE = S] (confidence: 0.968)
 [C233COMDPLTölüm = N, c53uçaktoplamaçuş = B] --> [C49KALİTE = S] (confidence: 0.969)
 [C233COMDPLTölüm = N, c100insanhatası = E] --> [C49KALİTE = S] (confidence: 0.969)
 [C233COMDPLTölüm = N] --> [C49KALİTE = S] (confidence: 0.971)
 [C49KALİTE = S, C41CPMDPLTSERTİFİKA = FA] --> [c55son90gün = B, C52UZMANLIKİŞ = O] (confidence: 0.982)
 [C233COMDPLTölüm = N, C41CPMDPLTSERTİFİKA = FA] --> [c55son90gün = B, C52UZMANLIKİŞ = O] (confidence: 0.982)
 [C49KALİTE = S, C233COMDPLTölüm = N, C41CPMDPLTSERTİFİKA = FA] --> [c55son90gün = B, C52UZMANLIKİŞ = O] (confidence: 0.982)
 [C41CPMDPLTSERTİFİKA = FA] --> [c55son90gün = B, C52UZMANLIKİŞ = O] (confidence: 0.983)
 [C49KALİTE = S, C41CPMDPLTSERTİFİKA = FA] --> [c55son90gün = B] (confidence: 0.983)
 [C49KALİTE = S, C52UZMANLIKİŞ = O, C41CPMDPLTSERTİFİKA = FA] --> [c55son90gün = B] (confidence: 0.983)
 [C233COMDPLTölüm = N, C41CPMDPLTSERTİFİKA = FA] --> [c55son90gün = B] (confidence: 0.983)
 [C49KALİTE = S, C233COMDPLTölüm = N, C41CPMDPLTSERTİFİKA = FA] --> [c55son90gün = B] (confidence: 0.983)
 [C233COMDPLTölüm = N, C52UZMANLIKİŞ = O, C41CPMDPLTSERTİFİKA = FA] --> [c55son90gün = B] (confidence: 0.984)
 [C49KALİTE = S, C233COMDPLTölüm = N, C52UZMANLIKİŞ = O, C41CPMDPLTSERTİFİKA = FA] --> [c55son90gün = B] (confidence: 0.984)
 [C41CPMDPLTSERTİFİKA = FA] --> [c55son90gün = B] (confidence: 0.984)
 [C52UZMANLIKİŞ = O, C41CPMDPLTSERTİFİKA = FA] --> [c55son90gün = B] (confidence: 0.984)
 [C233COMDPLTölüm = N, c47diğerrating = D] --> [C49KALİTE = S] (confidence: 0.985)
 [C49KALİTE = S, C41CPMDPLTSERTİFİKA = FA] --> [C52UZMANLIKİŞ = O] (confidence: 0.998)
 [C49KALİTE = S, C233COMDPLTölüm = N, C41CPMDPLTSERTİFİKA = FA] --> [C52UZMANLIKİŞ = O] (confidence: 0.998)
 [C41CPMDPLTSERTİFİKA = FA] --> [C52UZMANLIKİŞ = O] (confidence: 0.998)
 [C233COMDPLTölüm = N, C41CPMDPLTSERTİFİKA = FA] --> [C52UZMANLIKİŞ = O] (confidence: 0.998)
 [C49KALİTE = S, c55son90gün = B, C41CPMDPLTSERTİFİKA = FA] --> [C52UZMANLIKİŞ = O] (confidence: 0.998)
 [c53uçaktoplamaçuş = B, C41CPMDPLTSERTİFİKA = FA] --> [C52UZMANLIKİŞ = O] (confidence: 0.999)
 [c55son90gün = B, C41CPMDPLTSERTİFİKA = FA] --> [C52UZMANLIKİŞ = O] (confidence: 0.999)
 [C49KALİTE = S, C233COMDPLTölüm = N, c55son90gün = B, C41CPMDPLTSERTİFİKA = FA] --> [C52UZMANLIKİŞ = O] (confidence: 0.999)
 [C233COMDPLTölüm = N, c55son90gün = B, C41CPMDPLTSERTİFİKA = FA] --> [C52UZMANLIKİŞ = O] (confidence: 0.999)

Kaza

PredictiveApriori
=====

Best rules found:

1. c94kazakodutürü=GX C80YARDIMCINEDENFAKTÖRÜ=R 273 ==>
C184DESTEKLEYİCİOPERASYONELKOD=E 273 acc:(0.99438)
2. c94kazakodutürü=HB c98hasarkod=D C80YARDIMCINEDENFAKTÖRÜ=S 235 ==>
C184DESTEKLEYİCİOPERASYONELKOD=F 235 acc:(0.99406)
3. c78birincinedenfaktörü=IP c98hasarkod=D
C184DESTEKLEYİCİOPERASYONELKOD=F 230 ==> C80YARDIMCINEDENFAKTÖRÜ=S 230
acc:(0.994)
4. toplamölen=A c78birincinedenfaktörü=IP
C184DESTEKLEYİCİOPERASYONELKOD=F 219 ==> C80YARDIMCINEDENFAKTÖRÜ=S 219
acc:(0.99387)
5. C184DESTEKLEYİCİOPERASYONELKOD=F C80YARDIMCINEDENFAKTÖRÜ=T 200 ==>
toplamölen=A c98hasarkod=D 200 acc:(0.99359)
6. c94kazakodutürü=IA c78birincinedenfaktörü=IP
C80YARDIMCINEDENFAKTÖRÜ=S 194 ==> c98hasarkod=D 194 acc:(0.99349)
7. c94kazakodutürü=IA c78birincinedenfaktörü=IP
C184DESTEKLEYİCİOPERASYONELKOD=F 192 ==> c98hasarkod=D 192
acc:(0.99345)
8. toplamölen=A c94kazakodutürü=IA c78birincinedenfaktörü=IP 186 ==>
c98hasarkod=D 186 acc:(0.99334)
9. toplamölen=A c94kazakodutürü=IA c98hasarkod=D
C80YARDIMCINEDENFAKTÖRÜ=S 184 ==> c78birincinedenfaktörü=IP 184
acc:(0.9933)
10. c78birincinedenfaktörü=IP C184DESTEKLEYİCİOPERASYONELKOD=F 269 ==>
C80YARDIMCINEDENFAKTÖRÜ=S 268 acc:(0.9931)

Makine

PredictiveApriori
=====

Best rules found:

1. c31gövdeuçuşsaati=BA c157iniştakım=RC C27BÖLGESERTİFİKA=SO 275 ==>
c151motorsayısı=C 275 acc:(0.99495)
2. c157iniştakım=RC C155MOTORKOD=R C27BÖLGESERTİFİKA=SO 257 ==>
c151motorsayısı=C 257 acc:(0.99493)
3. c145ağırlık=OV c100insanhatası=H C27BÖLGESERTİFİKA=NM 206 ==>
c153beygirgücü=O c151motorsayısı=C 206 acc:(0.99487)
4. c145ağırlık=OV c153beygirgücü=O c100insanhatası=H
C27BÖLGESERTİFİKA=NM 206 ==> c157iniştakım=RT 206 acc:(0.99487)
5. c157iniştakım=RT C27BÖLGESERTİFİKA=EA 195 ==> c151motorsayısı=C
195 acc:(0.99485)
6. c145ağırlık=OV c31gövdeuçuşsaati=BC 181 ==> c153beygirgücü=O 181
acc:(0.99481)
7. c145ağırlık=OV C27BÖLGESERTİFİKA=EA 176 ==> c153beygirgücü=O
c151motorsayısı=C 176 acc:(0.9948)
8. c31gövdeuçuşsaati=BD C27BÖLGESERTİFİKA=NM 171 ==> c145ağırlık=OV
c153beygirgücü=O 171 acc:(0.99478)
9. c31gövdeuçuşsaati=BD C27BÖLGESERTİFİKA=NM 171 ==> c145ağırlık=OV
c151motorsayısı=C 171 acc:(0.99478)

10. c145ağırlık=OV c31gövdeuçuşsaati=BD C155MOTORKOD=T 169 ==>
c151motorsayısı=C 169 acc:(0.99477)

Ortam 1

PredictiveApriori
=====

Best rules found:

1. c117pistdurumu=ICE c100insanhatası=H 136 ==> c110ışıkdurumu=N 108
acc:(0.78362)
2. c75kazadakaçuçakvar=S 44 ==> c100insanhatası=H 31 acc:(0.69407)
3. c110ışıkdurumu=N c117pistdurumu=WTR 2 ==> c100insanhatası=H 2
acc:(0.68296)
4. c110ışıkdurumu=DU c100insanhatası=H 10 ==> c117pistdurumu=ICE 7
acc:(0.64373)
5. c110ışıkdurumu=DU 28 ==> c117pistdurumu=ICE 17 acc:(0.59832)
6. c110ışıkdurumu=DA 10 ==> c100insanhatası=H 5 acc:(0.48588)
7. c75kazadakaçuçakvar=F 45 ==> c100insanhatası=H 22 acc:(0.47952)
8. c110ışıkdurumu=N c117pistdurumu=WET 40 ==> c100insanhatası=H 17
acc:(0.4258)
9. c110ışıkdurumu=N c117pistdurumu=SNU 12 ==> c100insanhatası=H 5
acc:(0.42054)
10. c117pistdurumu=WTR c100insanhatası=H 5 ==> c110ışıkdurumu=N 2
acc:(0.42013)

Ortam 2

PredictiveApriori
=====

Best rules found:

1. c15havaalanıismi=WICHITA C106UÇUŞDURUM=C 20 ==> c100insanhatası=H
20 acc:(0.96306)
2. c15havaalanıismi=EPPLEY A C106UÇUŞDURUM=C 16 ==> c100insanhatası=H
16 acc:(0.95194)
3. c15havaalanıismi=KANSAS C C106UÇUŞDURUM=C 13 ==> c100insanhatası=H
13 acc:(0.93881)
4. c15havaalanıismi=SALT LAK C106UÇUŞDURUM=C 9 ==> c100insanhatası=H
9 acc:(0.90847)
5. c15havaalanıismi=DULLES I 7 ==> c100insanhatası=H 7
acc:(0.88187)
6. c15havaalanıismi=EPPLEY A C106UÇUŞDURUM=A 6 ==> c100insanhatası=H
6 acc:(0.86285)
7. c15havaalanıismi=CLEVELAN c100insanhatası=H 14 ==> C106UÇUŞDURUM=C
13 acc:(0.86199)
8. c15havaalanıismi=CLEVELAN C106UÇUŞDURUM=C 14 ==> c100insanhatası=H
13 acc:(0.86199)
9. c15havaalanıismi=GREATER C106UÇUŞDURUM=C 21 ==> c100insanhatası=H
19 acc:(0.85636)
10. c15havaalanıismi=GEORGE A 5 ==> c100insanhatası=H 5
acc:(0.8376)

ÖZGEÇMİŞ

Bursa’da 1980 yılında doğdu. İlk, orta, lise öğrenimini Bursa’da tamamladı. 1999 yılında Hava Harp Okulu’nu kazanarak Endüstri Mühendisliği Bölümü’nde lisans eğitimine başladı. 2003 yılında Teğmen olarak 2’nci Ana Jet Üs Eğitim Merkezi Komutanlığı’na atandı.

Mayıs 2004–Mayıs 2005 yılları arasında, Hava Teknik Okullar Komutanlığı Gaziemir/İZMİR Kursiyer Subay; Mayıs 2005-Temmuz 2007 8nci Ana Jet Üs Komutanlığı Levazım Şube Müdürlüğü DİYARBAKIR Giyim Teçhizat Subayı; Temmuz 2007-Ekim 2008 7. Kolordu Komutanlığı Orduevi Müdürlüğü Levazım Müdürü görevlerinde bulunduktan sonra Ekim 2008-Ekim 2010 Hava Harp Okulu Havacılık ve Uzay Teknolojileri Enstitüsü’nde Endüstri Mühendisliği Ana Bilim Dalı’nda Yüksek Lisans öğrenimini tamamlayarak 2010 atamalarıyla Hava Kuvvetleri Komutanlığı Karargahı Lojistik Başkanlığı Levazım Şube ANKARA’ya proje subayı olarak atanmıştır.

Evli ve İngilizce bilmektedir.