

TÜRKİYE'DE BİR HAVAYOLU İŞLETMESİNE AİT PARÇA SÖKÜM RAPORLARINA İLİŞKİN VERİ MADENCİLİĞİ UYGULAMASI

Feyza GÜRBÜZ, Lale ÖZBAKIR ve Hüseyin YAPICI*

Endüstri Müh. Bölümü, Erciyes Üniversitesi, 38039, Kayseri, feyza@erciyes.edu.tr, lozbakir@erciyes.edu.tr

* Makine Müh. Bölümü, Erciyes Üniversitesi, 38039, Kayseri, yapici@erciyes.edu.tr

(Geliş/Received: 19.03.2008 ; Kabul/Accepted: 04.11.2008)

ÖZET

Veri madenciliği, farklı bir çok alana başarı ile uygulanan bir tekniktir. Havacılık sektörü de bu alanlardan birisidir. Özellikle Türkiye'de gelişmekte olan hava yolu işletmelerinde de kullanılması mümkündür. Havacılık sektörü, çok büyük miktarlarda veri ve bilgi birikiminin olduğu bir alandır. Bu veriler pilot, bakım, kaza, parça söküm veya tehir raporları şeklinde kayıt altına alınmaktadır. Bu çalışmada, Türkiye'de bir hava yolu işletmesinin parça söküm raporları üzerinde veri madenciliği çalışması gerçekleştirilmiştir. Çalışmanın amacı, uçaklarda kullanılan parçaların, herhangi bir arıza oluşmadan önce düzeltici ve önleyici işlemlerin yapılması için ikaz seviyelerinin tespit edilmesine yönelik kural geliştirmektir. Sonuç olarak parçaların ikaz seviyelerini temsil edecek anlamlı bir kural elde edilmiş ve bulunan kurallar doğrulukları ve güvenilirlikleri bakımından test edilmiştir.

Anahtar Kelimeler: Veri madenciliği, havacılık, sınıflandırma, kural bulma.

DATA MINING APPLICATION ON COMPONENT REPORTS OF AN AIRLINE COMPANY IN TURKEY

ABSTRACT

Data mining methods have been successfully applied to different fields. Aviation industry is one of these fields. There is a large amount of knowledge and data accumulation in aviation industry. These data could be stored in the form of pilot reports, maintenance reports, incident reports, component reports or delay reports. This paper explains the data mining application on the component reports of an airline company in Turkey. In this study, we applied data mining functions on the component reports. The purpose of our application is to find out a rule about the warning level of the components to do correcting operation before any fault. As a result we found out a meaningful rule that presents the formulation of the warning level of the components. The rules found are tested in terms of their accuracy and reliability.

Keywords: Data mining, aviation, classification, find laws.

1. GİRİŞ (INTRODUCTION)

Son yıllarda hava taşımacılığında meydana gelen hızlı gelişmelerle uçuş ertelemeleri ve iptalleri ile birlikte kazalarda da hızlı bir artış gözlenmektedir [1].

Tipik bir havayolu yılda yapısal ve metinsel veri içeren, raporlar toplamaktadır. Kuruluşlar çeşitli kaynaklardan bu tip havacılık güvenlik bilgilerini toplayabilmek amacıyla önemli yatırımlar yapmaktadırlar. Bu bilgilerin depolandığı veritabanları birçok kritik veri içermektedir; bununla

birlikte bilginin analiz edilmesinde büyük problemler yaşanmaktadır. Havacılık endüstrisinde büyük hacimli karma yapıdaki verilerin zeki otomatik analizinde; yetenekli araçlara ihtiyaç duyulmaktadır. Bu araçlardan biri olan veri madenciliği, önceden bilinmeyen ve potansiyel olarak faydalı olabilecek, veri içinde gizli bilgilerin çıkarılmasına yarayan bir yöntemdir [2].

1960'lı yıllarda veri toplama ile başlamış olan veritabanı teknolojilerinin gelişim süreci, 1970'lerdeki veritabanları oluşturulması ile devam etmiştir.

1980’li yıllarda gelişen İlişkisel Veritabanı Yönetim Sistemi-RDBMS, ileri veri modelleri ve uygulama kaynaklı veritabanı yönetim sistemleri 1999 ve 2000’li yıllarda yerini veri madenciliği (VM) ve veri ambarlarına bırakmıştır [3].

Veritabanlarından Bilgi Keşfi ve Veri Madenciliği terimleri veritabanlarında aynı anlamı ifade eder. **Veritabanlarından bilgi keşfi**, veri içerisindeki geçerli, yeni, yararlı ve sonuç olarak anlaşılabilir örüntülerin çıkarılması sürecidir [4]. Veriler bir veritabanına kaydedilir. Bilgi tabanındaki bilgi keşfi sürecinden sağlanan bilgi basit hesaplarda elde edilemez. Bilgi kullanıcı için yeni ve avantajlı olmalıdır. Bazı doğruluk dereceleriyle yeni veriye uygulanabilmelidir. Son olarak bilgi insanın anlayabileceği şekilde olmalıdır. Bunun yanında veri madenciliği terimi verideki kullanışlı örneklerin bulunmasını ifade eder. Veri madenciliği, verilerden modeller veya örnekler üretmek için veri analizleri ve keşif algoritmaları kullanmayı gerektirir.

Veritabanındaki bilgi keşfi süreci, birkaç adımdan oluşan etkileşimli ve iteratif bir süreçtir. Bu süreç, uygulama alanının öğrenilerek, uygulamanın amaçları doğrultusunda hedef veri setinin seçilmesi ile başlar. Daha sonra, gürültülü, boş ve tutarsız verilerin çıkarıldığı veri temizleme ve ön işleme basamağı gelir. Gerekli durumlarda veri, madenciliğe uygun bir forma dönüştürülür. Sonraki aşama olan **veri madenciliği**, zeki yöntemler aracılığıyla büyük miktarda veriden anlamlı bilgilerin çıkarılması sürecidir. Daha sonra, çıkarılan örüntüler, içlerinden yararlı olanların belirlenmesi için değerlendirilir. Veritabanlarında bilgi keşfinin son basamağı ise, elde edilen bilginin görüntüleme ve bilgi gösterimi yöntemleri kullanılarak kullanıcıya sunulmasıdır. Veri madenciliği modellerini gördükleri işlemlere göre sınıflama (classification) ve regresyon (regression), kümeleme (clustering) ve birliktelik kuralları (association rules) olmak üzere üç ana başlık altında incelemek mümkündür [5]. Veri madenciliği, istatistik, makine öğrenimi, veri yönetimi ve veri tabanları, yapay zeka gibi alanların kesişimi esasına dayalı bir disiplindir [3].

Bu çalışmada, parça raporlarına veri madenciliği teknikleri polyanalyst adlı araç kullanılarak uygulanmıştır. Parçalara ait ikaz seviyelerinin belirlenmesine yönelik sınıflandırma ile kural keşfi süreci açıklanmıştır.

Parça ikaz seviyesi ile ilgili ilişkileri ve kuralları tespit etmek için lineer regresyon, yapay sinir ağı, bağımlılıkları bulma ve kural bulma algoritmaları kullanılmıştır.

Bu bölümde literatürde yer alan havacılık sektörüne yönelik veri madenciliği çalışmaları kısaca özetlenmiştir.

Bineid ve Fielding “Sivil yolcu uçaklarının güvenli uçuşunun geliştirilmesi için tahmin yöntemi” adlı çalışmalarında, yolcu uçaklarının sefere gönderilme güvenilirliklerinin tahmin edilmesini anlatmaktadır. Yöntem sonraki düzeltici işlemleri ve maliyetlerin en aza indirilmesi amacıyla korelasyon analizini kullanmışlardır. Tahmin yönteminin geliştirilmesi üç aşamada olmuştur: (1) Veri tabanı geliştirme, (2) Tahmin eşitliklerini türetme, (3) Modelin onaylanmasıdır [6]. Nazeri ve Zhang 2002’de Ulusal Havacılık Sistemi (NAS) performansını etkileyen ağır hava şartlarının analizi ile ilgili olan “Ağır Hava Şartlarının Hava Sahası Sistem Performansı Üzerindeki Etkilerini Anlamak İçin Havacılık Verilerinin Madenciliği” adlı çalışmalarında; veri hazırlama ve özellik seçimi (feature selection)’nin ne kadar önemli olduğunu göstermişlerdir. Çalışmada hava ve hava trafiği verileri olmak üzere iki tip veri kullanılmıştır. Bu verilere sınıflandırma, regresyon ve gruplama işlemleri uygulanmıştır. Anlamlı sınıflandırma kuralları üretilmiştir ve hava trafik günleri için oluşturulan grupların Ulusal Havacılık Sistemi performansı ile uygunluğu görülmüştür [1]. Shyur farklı insan hatalarından kaynaklanan kazalara ilişkin risk olasılıklarını modellemek için çalışmıştır. Çalışmalarında kullanılan verilerini, Taiwan Sivil Havacılık Kuruluşu (CAA) tarafından oluşturulan Havacılık Uçuş Güvenlik Yönetimi Bilgi Sistemleri (FSMIS) veri tabanından elde etmişlerdir [7]. Tseng ve arkadaşları pilot dikkatsizlikleri ve motor kazaları arasındaki ilişkileri keşfetmek amacıyla karar ağaçları ve yapay sinir ağları tekniklerini kullanmışlardır. Bu amaçla kullandıkları verileri 2000 ve 2003 yıllarına ait Ulusal Hava Yolları Trafik Güvenlik Kuruluşu, Ölümlü Kazalar Analiz Raporları Sistemi (FARS)’den almışlardır [8]. Solomon ve arkadaşları da ölümlü kazaları azaltmak amacıyla geliştirilen kırmızı ışık sinyalleri kontrol kameralarının kullanımının geliştirilmesinde veri madenciliği tekniklerinden karar ağaçları, yapay sinir ağları, sepet analizi, ve k-ortalamları tekniklerini kullanmışlardır [9].

2. VERİ MADENCİLİĞİ UYGULAMASI (DATA MINING APPLICATION)

2.1. Veri Kümesi (Dataset)

Çalışma için seçilmiş olan hava yolu şirketi ile yapılmış olan görüşmeler ve incelemeler sonucunda, veri madenciliği çalışmasının parça raporları üzerinde yapılabileceğine karar verilmiştir.

Parça raporlarında, parça sökümün planlı mı yoksa plansız mı olduğu, sökümün arızadan kaynaklanıp kaynaklanmadığı, parçanın uçaktaki miktarı gibi aşağıda belirtilen parametrelerle ilgili kayıtların yer aldığı tespit edilmiştir. Parça söküm raporlarında yer alan 31 parametre uzmanlarla birlikte yapılan değerlendirme sonucunda gereksiz ve boş bilgilerin çıkarıl-

masıyla 18 parametreye indirgenmiştir. Bu parametreler ve tanımları şöyledir:

- (1) Ata Chapter: Sırasıyla uçağa ait sistem (chapter/system), alt sistem (section/subsystem), parça (subject/unit), sayfa bloğu (page block)’nu tanımlar.
- (2) Thpyk: Hava yolu şirketinin her bir parçaya verdiği koddur.
- (3) Fleet: Parçanın hangi filoya ait olduğunu göstermektedir.
- (4) Description: Parçanın adını açıklamaktadır.
- (5) Qpa: Parçadan bir uçakta kaç tane olduğunu göstermektedir.
- (6) Hours: Parçanın kaç saat uçtuğunu belirtir.
- (7) Fs Total: Son 12 ayda belirtilen parçanın kaç kere söküldüğünü gösterir.
- (8) Fs01: Son 12 aydaki sökümlerden kaç tanesinin planlı yapıldığını belirtir.
- (9) Fs02: Son 12 aydaki sökümlerden kaç tanesinin plansız olarak yapıldığını belirtir.
- (10)Fs Other: Son 12 aydaki sökümlerin diğer söküm sebepleridir.
- (11)Fail: Son 12 aydaki sökümlerin kaç tanesinin arızalı olduğunu belirtir.
- (12)Not Fail: Son 12 aydaki plansız sökümlerden kaç tanesinin faal olduğunu gösterir.
- (13)Fail01: Son 12 aydaki planlı sökümlerin kaç tanesinin arızalı olduğunu gösterir.
- (14)Fail02: Son 12 aydaki plansız sökümlerin kaç tanesinin arızalı olduğunu gösterir.
- (15)Mtbr: Sökümler arasındaki ortalama süre.
- (16)Mtbur: Plansız sökümler arasındaki ortalama süre.
- (17)Mtbf: Arızalar arasındaki ortalama süre.
- (18)Falert: İkaz seviyesi değeri.

Analizin amacı sayısal bir parametre olan falert (parça ikaz seviyesi) parametresine etki eden diğer parametrelerin tespit edilerek, falert parametresinin tahmini için kural çıkarımıdır.

2.2. Metot (Method)

Veri madenciliği, büyük miktardaki veri içerisinde yararlı bilgiyi çıkarmayı amaçlamaktadır [10]. Kullanıcı için yararlı ve kullanışlı sonuçlar elde etmek için, veri tabanında başlangıçta bilinmeyen düzen ve ilişkilerin keşfedilmesi amacıyla, büyük miktardaki veri kümesinin seçimi, keşif ve araştırılması, ve modellenmesi sürecidir [11].

Veri madenciliğinde kullanılan modelleri, tahmin edici (Predictive) ve tanımlayıcı (Descriptive) olmak üzere iki ana başlık altında toplayabiliriz [12]. *Tahmin edici modellerde*; sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır. *Tanımlayıcı modellerde*; ise karar vermeye rehberlik etmede kullanılabilecek mevcut

verilerdeki örüntülerin tanımlanması sağlanmaktadır. Veri madenciliği, sınıflandırma, kümeleme, kural çıkarımı, benzerlik tespiti ve tanımlama gibi fonksiyonları yerine getirir [10]. Bu çalışmada bizim amacımız tahminleme amacıyla sınıflandırma işlemi gerçekleştirmektir.

2.2.1. Sınıflandırma (Classification)

En çok kullanılan veri madenciliği fonksiyonlarından birisi olan sınıflandırma işlemi, insan düşünce yapısına en yakın veri madenciliği işlemlerinden birisidir. Sınıflandırmada genellikle sınıflar arasındaki sınırların modellenmesiyle ilgilenilmektedir [13]. Bilinmeyen veri örneklerinin özelliklerinin incelenip, bu örneklerin daha önce tanımlanmış olan sınıflardan hangisine atanacağına karar verilmesi gerekmektedir. Burada sınıfların sayısı ve tanımlamaları daha önceden bilinmektedir. Sınıfı bilinmeyen yeni veri örneklerinin bu sınıflara atanması gerçekleştirilir. Sınıflandırma danışmanlı öğrenme olarak da bilinir. Daha çok tahminlemede kullanılır [14].

Polyanalyst programında yer alan sınıflandırma modülü veri kümesini temel iki gruba ayırır. Çıktı olarak bir sınıflandırma kuralı oluşturur. Kullanıcının seçeceği, kural bulma, polynet tahminleyicisi ve lineer regresyon alt algoritmalarına göre çalışır. Kural bulma algoritması büyük veri kümeleri için daha yavaş çalışır ancak daha güvenilir kurallar oluşturur. Polynet tahminleyicisi büyük veri kümeleri için daha hızlı çalışır. Lineer regresyon ise en hızlı çalışan modüldür. Ancak lineer olmayan bağımlılıklar için doğru modeller geliştiremeyebilir [15].

2.2.1.1. Lineer regresyon (Linear Regression)

Veriler arasındaki korelasyonu bulmak için en yaygın kullanılan yaklaşımlardan birisi lineer regresyondur [14]. Lineer regresyon istatistiksel tahminlemede en iyi bilinen ve en eski metotlardan birisidir. Lineer regresyon y değişkeninin x değişkenine bağımlılığını bulmak için x,y ekseninde bir doğru çizme prosesi olarak tanımlanabilir.

2.2.1.2. Kural bulma (Find Laws)

Bu algoritma veri kümesindeki doğrusal olmayan bağımlılıkları temsil etmek üzere sembolik kurallar oluşturmaktadır. Hedef değişkeni sayısal bir parametre olmalıdır. Çoklu regresyon mantığına dayanarak çalışır. Lineer regresyonun zayıf sonuçlar verdiği veriler üzerinde uygulanır. Çıktı olarak; en uygun kural ve en kesin kural olmak üzere iki tip kural üretir. En uygun kural, istatistiksel olarak uygun olan bir modeli temsil eder. En kesin kural ise en doğru modeli temsil eder. En uygun kural istatistiksel olarak güvenilirlik testlerinden geçmiş olan kuralken, en kesin kural tam olarak test edilmemektedir. Yani başka veri kümesi için tam doğru olmayabilir. Bu yüzden

uygulamalarda en uygun kuralın kullanılması gerektiği önerilmektedir. Kuralların doğruluğu ve etkinlikleri regresyon katsayısına (R^2) göre değerlendirilmektedir. R^2 değeri ($1-\rho^2$) değeri ile hesaplanır. Buradaki ρ değeri standart hatayı ifade etmektedir. R^2 değeri 0 ile 1 arasında her hangi bir sayıdır. Bu değer 1’e yaklaştıkça elde edilen kuralın doğruluğu artmaktadır [15]. Kural bulma, Megaputer Intelligence tarafından geliştirilmiş olan Symbolic Knowledge Acquisition Technology (SKAT), adlı algoritma esasına bağlı olarak çalışır. Bu algoritma, lineer olmayan bağımlılıkları temsil eden yüksek dereceli matematiksel ifadeler oluşturur. Yapay sinir ağlarının aksine Skat, çıktı olarak anlaşılabilir sembolik ifadeler verir.

2.2.1.3. PolyNet tahminleyicisi (PolyNet)

Polynet tahminleyicisi polyanalyst programının yapay sinir ağı aracıdır. Sayısal bir hedef değişkenin tahmininde kullanılır. Düğüm noktalarından oluşan bir yapay sinir ağı oluşturur. Her bir nokta diğer parametrelere bağlı olarak bir parametrenin tahmini için bir kural içerir. Eğer sinir ağı basitse kurallar sembolik olarak temsil edilebilir. Ancak çok karmaşık bir yapıya sahipse bir kara kutu gibi çalışır ve sembolik bir kural elde edilemez. Kural bulma algoritmasına göre daha fazla veri ile daha hızlı çalışma avantajına sahiptir. Ancak kategorik verilere uygulanmaz. Elde edilen kuralların doğruluğu ve etkinlikleri kural bulma modülündeki gibi regresyon katsayısına (R^2) göre değerlendirilmektedir. Bu katsayı 0 ile 1 arasında her hangi bir sayıdır ve 1’e yaklaştıkça elde edilen kuralın doğruluğu artmaktadır [15].

2.2.2. Bağımlılıkları bulma analizi (Find dependencies)

Bağımlılık modelleri değişkenler üzerindeki bağımlılıkları tanımlar. Yapısal ve ölçümsel olmak üzere iki seviyede gerçekleştirirler. Yapısal modeller yerel olarak hangi değişkenlerin bağımlı olduklarını belirler. Sayısal modeller ise sayısal ölçülerle bağımlılıkların kuvvetini gösterir. Bağımlılık modellerine örnek olarak olasılık grafikleri ve bulanık grafikler verilebilir. Diğer bağımlılık modeli yaklaşımlarına istatistiksel analizleri (korelasyon katsayısı, faktör analizi gibi) ve yapay sinir ağları kullanılarak yapılan duyarlılık analizleri örnek gösterilebilir [14.].

Polyanalyst programında yer alan bağımlılıkları bulma algoritması genel olarak ön işleme aracı olarak tasarlanmış, açıklanabilir kuralları olmayan, muhtemel en kuvvetli bağıntıları veya aykırılıkları tespit eden bir algoritmadır. Veri kümesi içerisindeki

ilişkileri ve bağımlılıkları bulmak için kullanılır. Değişkenler arasındaki bağlantının ne kadar güçlü olduğunu gösterir. Bağımlılıkları bulma ile veri kümesinde bulunan kötü veriler elimine edilebilir. Hedef değişkeni sayısal veya ikili (binary) bir parametre olmalıdır. Çıktı olarak, bir tahmin kuralı ile keşfedilen bağımlılıklara uyan veri sayısını vermektedir. PolyAnalyst bağımlılıkları bulma modülü Arnavac algoritmayı kullanır. Bu algoritma veri kümesi içerisindeki karışık parametrelerin, hedef değişken üzerindeki etkisini ölçme esasına dayanır. Bağımlılık ölçüsü olarak analiz sonucunda elde edilen p değerlerine bakılır. Bu değer [0 1] arasında bir değer alır. P değeri sıfıra ne kadar yakın olursa değişkenler arasındaki bağımlılığın şiddeti o kadar yüksek olur. P değeri 10^{-7} değerinin altında ise 0 (sıfır) kabul edilir [15].

2.3. Veri Madenciliği (Data Mining)

Analiz sürecimizin ilk adımı verilerin ön işlemlerinin yapılarak sınıflandırma algoritması uygulanacak verinin oluşturulmasıdır. Gerekli ön işlemlerle ikaz seviyesinin belirlenmesinde etkili parametreler tespit edildikten sonra mevcut veri kümesi test ve eğitim verileri olmak üzere ikiye ayrılmıştır. Veri kümesinin %30’u test verisi, geriye kalan %70’i ise eğitim verisi olarak program yardımıyla rasgele oluşturulmuştur. Eğitim verileri ile analiz gerçekleştirilmiş; test verileri üzerinde de bulunan kuralın doğruluğu ve güvenilirliği test edilmiştir.

2.3.1. Ön işleme (Preprocessing)

Ön işleme sürecine daha önce tanımları yapılan 18 parametre olan veri kümesi, {ata chapter, component kodu, fleet, description, qpa, hours, fstotal, fs01, fs02, fs0ther, fail, notfail, fail01, fail02, mtr, mtrbur, mtrbf, falert}, ile başlanmıştır. İlk olarak bütün bu parametreler arasında birbirleriyle bağlantılı olan parametreler regresyon analizi ile bulunmuştur. Regresyon analizi sonuçlarına göre fs total (toplam söküm), parametresi ile fs01, fs02 ve fs other parametreleri arasında; fail(toplam arıza) parametresi ile de fail01 ve fail02 parametreleri arasında ilişki olduğu tespit edilmiştir. Dolayısıyla fs total parametresi, fs01, fs02 ve fs other parametreleri ile, fail parametresi ise fail01 ve fail02 parametreleri ile temsil edilebilir. Bundan dolayı ön işleme sürecinin ilk adımında belirtilen 18 parametreden, fs total ve fail parametrelerinin çıkarılmasına karar verilmiştir. Regresyon analizi sonuçları Tablo 1(a) ve Tablo 1(b)’de gösterilmiştir.

Regresyon analizinin ardından, geriye kalan 16 parametreden oluşan veri kümesi, {ata chapter, component kodu, fleet, description, qpa, hours, fs01, fs02,

Tablo 1(a). Regresyon analizi sonuçları (Results of the regression analysis)

<p>The regression equation is FS TOTAL-TOPLAM SÖKÜM = - 0.000000 + 1.00 PLANLI SÖKÜM-FS01 + 1.00 PLANSIZ SÖKÜM-FS02 + 1.00 DİĞ ER SÖKÜMLER-FSOTHER S = 0 R-Sq = 100.0% R-Sq(adj) = 100.0% PRESS = 6.460926E-20 R-Sq(pred) = 100.00%</p>

Tablo 1 (b). Regresyon analizi sonuçları (Results of the regression analysis)

The regression equation is:
 TOPLAM ARIZA-FAIL= 0.385 + 1.09 PLANSIZ ARIZA-FAIL02 + 0.732 FAIL01-PLANLI ARIZA
 S = 4.51563 R-Sq = 91.6% R-Sq(adj) = 91.6%
 PRESS = 1020284 R-Sq(pred) = 91.23%

fsother, notfail, fail01, fail02, mtbr, mtbur, mtbf, falert}, ile ön işleme sürecine devam edilmiştir. Bu veri kümesi ile bağımlılıkları bulma analizi yapılarak, ikaz seviyesine etkileyen parametreler tespit edilmiştir. Bağımlılıkları bulma analizine ait sonuçlar ve ilgili p değerleri Tablo 2’de gösterilmiştir. Etkili olan parametrelerin tespit edilmesinde kullanılan p değeri 0’a yaklaştıkça parametrenin hedef parametre üzerindeki etkisinin de arttığı kabul edilmektedir [15].

Bağımlılıkları bulma analizi sonucunda ikaz seviyesine etkisi olan 10 parametre, {Mtbf, Fail 02, Mtbur, Qpa, Hours, Fs 02, Not Fail, Mtbr, Fs 01, Mtbf-3} tespit edilmiştir. Tespit edilen etkili parametreler ve ikaz seviyesi sayısal değerler içeren parametrelerdir.

Tablo 2. Bağımlılıkları bulma analizi sonuçları (Results of the find dependencies analysis)

Parametre Adı	P değeri	Parametre Adı	P değeri
Mtbf	2.062 e^{-70}	Fs 02	1.088 e^{-19}
Fail 02	1.551 e^{-39}	Not Fail	1.088 e^{-19}
Mtbur	1.551 e^{-39}	Mtbr	1.088 e^{-19}
Qpa	5.902 e^{-24}	Fs 01	4.855 e^{-7}
Hours	5.902 e^{-24}	Mtbf-3	4.855 e^{-7}

2.3.2. Sınıflandırma analizi (Classification)

Sınıflandırma analizi ile, hava yolu şirketinin kullandığı Şekil 1’de gösterilen ikaz seviyelerinin tespitinde kullandıkları formüle alternatif olarak yeni bir kural oluşturmaktır. Gerçek bir ikazın görülmesi veya sahte ikaz olasılığı, k katsayısına bağlıdır. k genellikle 2 seçilir. Fakat arıza oranındaki yüksek

$$\text{Ikaz Seviyesi} = \bar{Y} \pm k\sigma$$

$$\sigma = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n}} = \text{standart sapma}$$

n= hesaplarda kullanılan ay sayısı,
 Y= aylık oran,
 \bar{Y} = Son n ay için aylık ortalama

Şekil 1. Havayolu şirketinin ikaz seviyelerinin tespitinde kullandıkları formülasyon (The formulation of warning level that the airline company uses.)

farklılıklar gösteren sistemler için meydana gelebilecek sahte ikazların sayısını azaltmak amacıyla k katsayısı 3 olarak seçilebilir. k katsayısı, mevcut çıkan ikaz değerlerine bağlı olarak 1 ile 3 arasında ayarlanabilir.

Ön işleme sonucunda elde edilen veri kümesi, {Mtbf, Fail 02, Mtbur, Qpa, Hours, Fs 02, Not Fail, Mtbr, Fs 01, Mtbf-3}, ile sınıflandırma analizleri gerçekleştirilmiştir. Polyanalyst aracının sınıflandırma algoritmaları olan, bağımlılıkları bulma, polynet tahminleyicisi ve lineer regresyon analizleri ayrı ayrı veri kümesi üzerine uygulanmıştır. Analiz sonuçlarında elde edilen sınıflandırma kuralları Tablo 3’de verilmiştir.

Elde edilen kuralların doğrulukları için R^2 değerlerine bakılmıştır. Bu değer [0-1] değerleri arasında bir değer almaktadır. R^2 değeri 1’e yaklaştıkça elde edilen kuralın doğruluğu da artmaktadır. Analiz sonucunda elde edilen test ve eğitim sonuçlarına ait R^2 değerleri Tablo 4’de gösterilmiştir.

Ortaya çıkan R^2 değerlerini karşılaştıracak olursak en doğru ve güvenilir olan kuralın, kural bulma algoritması ile elde edildiği görülmüştür. Dolayısıyla ikaz seviyelerinin tespit edilmesinde kullanabilmeleri için hava yolu şirketine önereceğimiz kural, kural bulma algoritması ile elde ettiğimiz tahmin kuralı olmuştur. Bu kurala göre parçaların ikaz seviyesi, uçaktaki miktarı, sökümler arasındaki ortalama süre ve arızalar arasındaki ortalama sürenin bir fonksiyonu olarak bulunmuştur.

2.3.3. Sonuçlar (Results)

Eğitim ve test verilerine ait analiz sonuçlarına göre elde edilen doğruluk değerlerini karşılaştırdığımızda, analizimizde kullandığımız tekniklerden bağımlılıkları bulma ve kural bulma algoritmaları ile daha güvenilir sonuçlar elde edildiği görülmüştür. Veri içerisindeki anlamlı, doğru ve kullanılabilir bilgiyi elde etmede bu iki algoritmanın kullanılmasının uygun olduğunu görmekteyiz. Elde ettiğimiz tahmin kuralı ile hava yolu şirketinin ilgili birim çalışanları her bir parça için ikaz seviyelerini kolayca tahmin ederek, olası arızalardan önce gerekli düzeltici ve önleyici bakım işlemlerinin yapılması mümkün olacaktır.

3. SONUÇLAR VE TARTIŞMA (CONCLUSIONS AND DISCUSSION)

Bu çalışmada havacılık sektöründe parça raporlarında farklı veri madenciliği tekniklerinin kullanılabileceğini göstermiş bulunmaktayız. İlk olarak verilerin temizlenmesi ve girdi ile çıktı parametreleri arasındaki bağımlılıkları tespit ederek ön işleme teknikleri uygulanmıştır. İkinci adımda havacılık parça verile-

Tablo 3. Sınıflandırma analizi sonucunda elde edilen tahmin kuralları (The prediction rules of classification analysis)

Kural Bulma ile elde edilen tahmin kuralı	$\text{Falert} = (0.656074 * \text{Mtbf} - 8.22693e^{-7} * \text{Mtbf}^2 + 6.34504e^{-10} * \text{Qpa}^2 * \text{Mtbf}^2 - 2.24397e^{-8} * \text{Mtb}r * \text{Qpa} * \text{Mtbf}) / (1 + 0.0580449 * \text{Qpa} - 1.81272e^{-6} * \text{Mtbf})$
Polynet ile elde edilen tahmin kuralı	$\text{Falert} = (15804.9 + \text{Hours} * (\text{Hours} * (7.81226e^{-14} * \text{Hours})) + \text{Mtb}r * (\text{Mtb}r * (-1.44442e^{-6} + 1.78044e^{-12} * \text{Mtb}r + 9.22241e^{-12} * \text{Hours}) + \text{Hours} * (1.09508e^{-5} - 1.57257e^{-11} * \text{Hours})))$
Lineer Regresyon ile elde edilen tahmin kuralı	$\text{Falert} = (-812.789 * \text{Qpa} - 799.893 * \text{Fs}01 - 1231.41 * \text{Notfail} + 547.711 * \text{Fail}02 - 0.196730 * \text{Mtb}r - 0.353921 * \text{Mtb}ur + 0.974162 * \text{Mtbf})$

Tablo 4. Sınıflandırma analizi sonucunda elde edilen R² değerleri (R² values of the classification results)

	Kural Bulma		Polynet		Lineer Regresyon	
	Eğitim	Test	Eğitim	Test	Eğitim	Test
R ²	0.9001	0.681894	0.3784	0.26789	0.7594	0.661696

rinden yararlı bilginin çıkarılması için sınıflandırma algoritmaları uygulanarak veri madenciliği kavram ve bileşenlerinden bahsedilmiştir.

Yapmış olduğumuz çalışma ile, Türkiye’de yaygınlaşmakta olan hava taşımacılığının geliştirilmesinde veri madenciliği tekniklerinin kullanılabilirliği ortaya konulmuştur. Mevcut veri tabanlarında yer alan farklı birçok veri ile yapılacak veri madenciliği çalışmaları ile kullanılabilir ve yararlı bilgiler elde edilebilir.

KAYNAKLAR (REFERENCES)

1. Nazeri, Z., Jianping, Z., Mining aviation data to understand impacts of severe weather on airspace system performance, **IEEE, Proceedings of the International Conference on Information Technology**, 2002.
2. Jiawei H., Kamber M., Data Mining: Concepts and Techniques, University of Simon Fraser, 2001.
3. Karakaş, M., Veri Madenciliği Üzerine, <http://www.bilgiyonetimi.org>.
4. Wong, M. L., Leung, K. S., Data Mining Using Grammar Based Genetic Programming & Applications, **Genetic Programming**, Vol.3, 2000, ISBN: 0-7923-7746-X.
5. Sever, H., Oğuz, B., Veri Tabanlarında Bilgi Keşfine Formal Bir Yaklaşım Kısım 1: Eşleştirme Sorguları ve Algoritmalar, <http://www.baskent.edu.tr/~sever/bilgi-p01-03.pdf>.
6. Bineid, M., Fielding, J. P., Development of a civil aircraft dispatch reliability prediction methodology, **Aircraft Engineering and Aerospace Technology**, Volume 75, Number 6, pp. 588-594, 2003.
7. Shyur, H. J., A quantitative model for aviation safety risk assesment, **Computers&Industrial Engineering**, 2007.
8. Tseng, W. S., Nguyen, H., Liebowitz, J., Agresti, W., Distractions and motor vehicle accidents:Data mining application on fatality analysis reporting system (FARS) data files, **Industrial Management & Data Systems**, Vol. 105, No. 9, 2005, pp 1188-1205.
9. Solomon, S., Nguyen, H., Liebowitz, J., Agresti, W., Using data mining to improve traffic safety programs, **Industrial Management & Data Systems**, Vol. 106 No. 5, 2006, pp. 621-643.
10. Emekci, F., Sahin, O. D., Agrawal, D., El Abbadi, Privacy preserving decision tree learning over multiple parties, **Data & Knowledge Engineering**, 63, 348-361, 2007.
11. Giudici, P., Applied data mining: statistical methods for business and industry, Faculty of Economics, University of Pavia, Italy, Wiley Publications, 2003.
12. Eker, H., Veri Madenciliği veya Bilgi Keşfi, http://www.bilgiyonetimi.org/cm/pages/mkl_gos.php?nt=538.
13. Hand, D., Manila, H., Smyth, P., Principles of Data Mining, Cambridge, Massachusetts, London, England, 2001.
14. Zhang, C., Zhang, S., Association Rule Mining, Models and Algorithms, Springer, 2002.
15. User Manuel of PolyAnalyst 5, April 2005.