
UÇAK SEFERLERİNDEKİ RÖTARLARI ETKİLEYEN FAKTÖRLERİN ANALİZİ

Aslı ÇALIŞ BOYACI¹

Korel İnanç DURMAZ²

Cevriye GENCER³

Öz

Her geçen yıl talebin daha da arttığı hava taşımacılığında, kapasite kısıtına bağlı gecikmeler en önemli konu haline gelmiştir. Çünkü yaşanan bu gecikmeler, havayolu şirketlerinin maliyetlerinde artışa sebep olurken müşteri memnuniyetinde de azalmaya neden olmaktadır. Özellikle artan trafik hacmi, havalimanlarında yaşanan bu gecikmelerin süresini uzatmaktadır. Trafik yoğunluğunun yanı sıra güney hâkim yönlü rüzgârlar (Iodos) kaynaklı rötaların eksik olmadığı Atatürk Havalimanı'nda yolcuların ve havayolu şirketlerinin yaşanan gecikmeler yüzünden büyük sıkıntı ve zarara uğradığı bilinen bir gerçektir. Pistte kalkış için sıraya giren ve iniş için bekleme noktasında bulunan uçakların meydana getirdiği kuyrukların Iodoslu havalarda gittikçe uzaması sonucu çok sayıda sefer gecikmekte ya da iptal edilmektedir. Güney yönlü rüzgârlar altında kapasitesinde azalma meydana gelen Atatürk Havalimanı kapasitesini arttırmak amacıyla gerçekleştirilen bu çalışmada, Veri Madenciliği ile uçak seferlerindeki rötaları etkileyen faktörler analiz edilmektedir.

Anahtar Kelimeler: Havalimanı, Rötâ, Veri Madenciliği

JEL Sınıflandırması: R41,R42

ANALYSIS OF THE FACTORS AFFECTING THE FLIGHT DELAYS

Abstract

In air transport, where demand is increasing every year, delays due to capacity constraints have become the most important issue. Because these delays cause increases in the costs of airline companies, they also cause a decrease in customer satisfaction. Increasing traffic volume, in particular, prolongs the delays experienced at airports. It is a fact that at the Atatürk Airport, where traffic delays are not lacking due to southwester, it is a known fact that passengers and airline companies suffer from great difficulties and damages due to the delays experienced. Queues, occurred by airplanes entering the queue for departure and waiting for landing, are getting lengthened during southwester and many expeditions are delayed or canceled. In this study, which aims at increasing the capacity of Atatürk Airport, which is a decrease in capacity under southwester, factors affecting the flight delays are analyzed with data mining.

Keywords: Airport, Retard, Data Mining

JEL Classification: R41,R42

¹ Yrd. Doç. Dr., Ondokuz Mayıs Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, asli.calis@omu.edu.tr

² Doktora Öğrencisi, Gazi Üniversitesi, Bilişim Enstitüsü, kidurmaz@gmail.com

³ Prof. Dr., Gazi Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, ctemel@gazi.edu.tr

1. Giriş

Tüm dünyada her geçen yıl artan hava trafiği hacmi, sınırlı kapasite kaynaklarına sahip olan havalimanları tarafından karşılanmaya çalışılmaktadır. Havalimanı kapasitesinin sınırlı olması ve hava trafik hacminin artması sonucu havalimanlarında gecikmeler meydana gelmektedir. Pistte kalkış için sıraya giren ve iniş için bekleme noktasında bulunan uçakların meydana getirdiği kuyrukların lodoslu havalarda gittikçe uzaması sonucu çok sayıda sefer iptal edilmekte ya da gecikmektedir. Çünkü lodoslu havalarda ve rüzgâr şiddetinin 4 Knot'dan büyük olduğunda Atatürk Havalimanı'nın V şekilli pist yapısından dolayı bir uçak kalkış yapmadan bir başka uçağın inişine ya da bir uçak iniş yapmadan diğer uçağın kalkışına izin verilmemektedir. Bu durumda uçaklarının havada attığı fazladan turlar, maddi kayıplara yol açtığı gibi uçakların karbondioksit emisyonunu arttırmakta olup, çevreyi de olumsuz etkilemektedir. Havalimanı kapasitesinin sınırlı olması, hava trafik hacim yoğunluğunun artması ve güney hâkim yönlü rüzgârlar (lodos), havayolu şirketlerinin maliyetlerinde artışa sebep olurken, hava kalitesinde ve müşteri memnuniyetinde azalmaya yol açmaktadır. Havalimanı kapasitesini uzun vadede arttırmanın yolu yeni bir pist veya yüksek kapasiteli havalimanı yapımıdır. Havalimanı kapasitesini kısa ve orta vadede arttırmanın yolu ise uluslararası havacılık kurallarına uygun olarak hava trafik kurallarında yapılacak değişikliklerdir. Bu çalışmada güney yönlü rüzgârlar (lodos) altında kapasitesinde azalma meydana gelen Atatürk Havalimanı kapasitesini arttırmak amacıyla uçak seferlerinde yaşanan gecikmeleri etkileyen faktörlerin analizi amaçlanmaktadır.

2. Literatür Özeti

Literatürde veri madenciliği üzerine gerçekleştirilmiş çok sayıda çalışma mevcuttur. Bu çalışmalardan bazıları şöyledir: Oğuzlar, hanehalkı işgücü anketi sonuçlarını CART algoritması kullanarak özetlemeye çalışmıştır (Oğuzlar, 2004: 79). Chien ve Chen, personel seçimi için karar ağaçları ve birliktelik kurallarına dayalı bir veri madenciliği uygulaması gerçekleştirmişlerdir (Chien ve Chen, 2008: 280). Taşkın ve Emel'e ait çalışmada, bir perakende işletmenin müşterilerinin Kohonen ağları ile kümelmesi ele alınmıştır (Taşkın ve Emel, 2010: 395). Dolgun ve Ersel karar ağaçları, lojistik regresyon, bayesci ağlar ve destek vektör makineleri gibi veri madenciliği yöntemlerini kullanarak doğrudan pazarlama kampanyalarının nasıl yönlendirilebileceği üzerinde durulmuşlardır (Dolgun ve Ersel, 2014: 1-13). Duran ve arkadaşları, bir bankanın doğrudan pazarlama kampanyasına katılabilecek hedef müşteri kitlesini belirleyebilmek için yapay sinir ağları, lojistik regresyon ve karar ağaçlarının tahmin güçlerini karşılaştırarak kampanyayı en iyi açıklayan modeli belirlemişlerdir (Duran vd., 2014: 142). Siknun ve Sitanggang, orman yangınlarına ilişkin veri seti üzerinde C5.0 algoritmasını kullanarak web tabanlı bir uygulama gerçekleştirmişlerdir (Siknun ve Sitanggang, 2016: 332).

Literatürde veri madenciliği ile ilgili birçok çalışma mevcut olmasına rağmen, havalimanı ve veri madenciliğinin birlikte ele alındığı sınırlı sayıda çalışmaya rastlanılmaktadır. Liou ve diğerlerine ait çalışmada, yolcuların havalimanı hizmet düzeyi algıları için geleneksel istatistiksel analizden farklı olarak, veri madenciliğinde sınıflandırma problemleri için kullanılan kaba küme yaklaşımı uygulanmıştır. Önerilen yöntem, havaalanı planlayıcılarına, tasarımcılara, operatörlere ve yöneticilere, hizmet düzeyi (LOS) iyileştirme stratejileri geliştirmeleri için gereken pratik bilgileri sağlamaktadır. Model, Tayvan'daki uluslararası bir havalimanından gelen geniş bir müşteri örneğinden alınan anket verileri kullanılarak uygulanmıştır (Liou vd., 2011: 13723). Fan ve diğerlerine ait çalışmada, veri madenciliğinde sınıflandırma yöntemlerinden karar ağacı ile uluslararası bir havalimanında buz çözme faaliyetlerinin çevresel etkileri üzerine bir istatistiksel analiz sunulmuştur (Fan vd., 2011: 14899). Smith ve Ehmke'ye ait çalışmada bir havalimanında hava yolu operasyonlarını, bir havalimanı ve bir tren istasyonu arasındaki taksi hizmetini ve bir havalimanından taksi hizmetlerini içeren üç problem türü için bir tamsayı programlama modeli geliştirilmiştir. Önerilen tamsayı programlama modelinin bilgi keşfi sürecindeki veri eşleştirme ve veri temizleme aşamaları için işlenebilir olduğu kanıtlanmıştır (Smith ve Ehmke, 2016: 375).

3. Veri Madenciliği

Veri madenciliği, tek başına ham verinin sunmadığı bilgiyi ortaya çıkaran veri analizi sürecidir (Jacobs, 1999: 8-9). Farklı bir ifadeyle veri madenciliği, büyük miktardaki veriden, anlamlı örüntüler ve kurallar keşfetme sürecidir (Linoff ve Berry, 2011). Veri madenciliğinin ortaya çıkışı veri yığınlarının geniş yer kaplamasına ve büyük miktardaki verilerin yararlı bilgilere dönüştürülmesi ihtiyacına dayanmaktadır (Han ve Kamber, 2006).

Veri madenciliği, kavramsal olarak 1960'lı yıllarda, bilgisayarların veri analiz problemlerini çözmek için kullanılmaya başlanmasıyla ortaya çıkmıştır. Bu dönemlerde veri taraması, veri yakalanması gibi isimler verilmiş ve bilgisayar yardımıyla gerekli sorgulama yapıldığında istenilen bilginin elde edilebileceği düşünülmüştür (Köktürk vd., 2009: 21). 1970'lerde ilişkisel Veri Tabanı Yönetim Sistemleri uygulamaları kullanılmaya başlanmıştır. Bilgisayar uzmanları bununla beraber basit kurallara dayanan uzman sistemler geliştirmişler ve basit anlamda makine öğrenimini sağlamışlardır. 1980'lerde veri tabanı yönetim sistemleri yaygınlaşmış ve bilimsel alanlarda, mühendisliklerde vb. alanlarda uygulanmaya başlanmıştır. Bu yıllarda şirketler, müşterileri, rakipleri ve ürünleri ile ilgili verilerden oluşan veri tabanları oluşturmuşlardır. Bu veri tabanlarının içerisinde çok büyük miktarlarda veri bulunmaktadır ve bunlara SQL veri tabanı sorgulama dili ya da benzeri diller kullanarak ulaşılabilir (Savaş vd., 2012: 4-5). 1990'larda bilgisayar mühendisleri, geleneksel istatistiksel yöntemlerinin yerine algoritmik bilgisayar modülleri ile veri analizinin değerlendirilebileceğini vurgulayarak, veri madenciliği ismini kullanmışlardır. Bu yıllarda veri tabanlarındaki veri miktarları katlanarak arttığı için, büyük miktardaki veri içinden yararlı bilgilere nasıl ulaşılması gerektiği üzerinde düşünölmeye başlanmıştır ve veri madenciliği için ilk yazılım gerçekleştirilmiştir. 2000'li yıllardan itibaren veri madenciliği sürekli gelişmiş ve geniş bir yelpazede uygulanmaya başlanmıştır.

Büyük hacimde veri bulunan her yerde veri madenciliğini kullanmak mümkündür. Veri madenciliği, makine öğrenme, istatistik ve veri tabanları alanlarındaki teknikleri birleştirerek, büyük veri tabanlarından faydalı ve değerli bilgiyi çıkarmamıza imkân tanımaktadır (Ching ve Pong, 2002). Veri madenciliği, pazarlama, finans, bankacılık, üretim, sağlık, müşteri ilişkileri yönetimi ve organizasyon öğrenme gibi karar verme sürecine ihtiyaç duyulan birçok alanda uygulanabilir olması nedeniyle son zamanlarda veri tabanı kullanıcıları ve araştırmacıların önemli ölçüde dikkatini çekmektedir. (Chien ve Chen, 2008: 280-290).

Veri madenciliği, aynı zamanda bir süreçtir. Veri yığınları arasında, soyut kazılar yaparak veriyi ortaya çıkarmanın yanı sıra, bilgi keşfi sürecinde örüntüleri ayrıştırarak bir sonraki adıma hazır hale getirmek de bu sürecin bir parçasıdır. Üzerinde inceleme yapılan işin ve verilerin özelliklerinin bilinmemesi durumunda ne kadar etkin olursa olsun hiç bir veri madenciliği algoritmasının fayda sağlaması mümkün değildir. Bu nedenle, veri madenciliği sürecine girilmeden önce, başarının ilk şartı, iş ve veri özelliklerinin detaylı analiz edilmesidir. Veri madenciliği sürecinde izlenen adımlar genellikle aşağıdaki şekildedir (Savaş vd., 2012: 7-8):

- **Problemin tanımlanması:** Veri madenciliği çalışmalarında başarılı olmanın en önemli şartı, projenin hangi amaç için yapılacağı ve elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceğinin tanımlanmasıdır.

- **Verilerin hazırlanması:** Modelin kurulması aşamasında ortaya çıkacak sorunlar, bu aşamaya sık sık geri dönülmesine ve verilerin yeniden düzenlenmesine yol açacaktır. Bu durum verilerin hazırlanması ve modelin kurulması aşamaları için, bir analistin veri keşfi sürecinin toplamı içerisinde enerji ve zamanının %50 - %85'ini harcamasına neden olmaktadır. Verilerin hazırlanması, "toplama", "değer biçme", "birleştirme ve temizleme", "örneklem seçimi" ve "dönüştürme" aşamalarından oluşmaktadır.

- **Modelin kurulması ve değerlendirilmesi:** Tanımlanan problem için en uygun modelin belirlenebilmesi, mümkün olduğunca çok sayıda modelin kurularak denenmesi ile mümkündür. Bu

nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar tekrarlanan bir süreçtir.

• **Modelin kullanılması:** Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir.

• **Modelin izlenmesi:** Zaman içerisinde bütün sistemlerde ve ürettikleri verilerde meydana gelen değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve yeniden düzenlenmesini gerektirecektir.

Veri madenciliğinde kullanılan modeller, *tahmin edici* ve *tanımlayıcı* olmak üzere iki ana başlık altında incelenmektedir. 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 kullanılacak mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır (Akpınar, 2000: 5).

Veri madenciliği modellerini işlevlerine göre üç ana grup altında toplamak mümkündür:

1. Sınıflama ve Regresyon
2. Kümeleme
3. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler

Sınıflama ve regresyon, önemli veri sınıflarını ortaya koyan veya gelecek veri eğilimlerini tahmin eden modelleri kurabilen analiz yöntemleridir. Sınıflama kategorik değerleri tahmin ederken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılmaktadır. Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler; Yapay Sinir Ağları, Genetik Algoritmalar, K- En Yakın Komşu, Naive-Bayes Sınıflayıcısı, Lojistik Regresyon ve Karar Ağaçları' dır. (Özekes ve Çamurcu, 2002: 1-17).

Kümeleme modellerinde amaç, küme üyelerinin birbirlerine çok benzediği ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere bölünmesidir (Akpınar, 2000: 6). Kümeleme analizi, önceden belirlenen seçme kriterlerine göre birbirine çok benzeyen birey ya da nesneleri aynı küme içinde gruplandırmaktadır. Analizin sonucunda bir kümeyi oluşturan birey veya nesneler birbirlerine benzerken, diğer kümelerin birey veya nesnelere benzemeyeceğinden, kümelerin kendi içlerinde homojenlik, kümeler arasında ise heterojenlik söz konusu olmaktadır. Oluşturulan kümeler çok boyutlu uzayda gösterildiğinde, eğer kümeleme başarılı ise aynı küme içinde yer alan birey veya nesnelerin birbirlerine oldukça yakın çıkması, bununla birlikte farklı kümelerin de birbirinden fark edilir düzeyde uzak olması beklenmektedir (Suner ve Çelikoğlu, 2010: 46).

Birliktelik kuralı, geçmiş verilerin analiz edilerek bu veriler içindeki birliktelik davranışlarının tespiti ile geleceğe yönelik çalışmalar yapılmasını destekleyen bir yaklaşımdır (Özçakır ve Çamurcu, 2007: 22). Satın alma eğilimlerinin tanımlanmasını sağlayan birliktelik kuralları ve ardışık zamanlı örüntüler, pazarlama amaçlı olarak "pazar sepeti analizi" adı altında veri madenciliğinde yaygın olarak kullanılmaktadır. Bununla birlikte bu teknikler, tıp, finans ve farklı olayların birbirleri ile ilişkili olduğunun belirlenmesi sonucunda değerli bilgi kazanımının söz konusu olduğu ortamlarda da önem taşımaktadır (Akpınar, 2000: 7).

Sınıflandırma, en yaygın olarak kullanılan veri madenciliği yöntemidir. Amacı, sınıfı belirlenmiş verilerden yola çıkılarak sınıfı belli olmayan verilerin, hangi sınıfa dâhil edileceğinin tahmin edilmesidir. Çalışmada sınıflandırma yöntemlerinden karar ağaçları ile uçak seferlerinde yaşanan rötaları etkileyen faktörler analiz edilmektedir. Bu doğrultuda bir sonraki aşamada karar ağaçlarına yer verilmiştir.

3.1. Karar Ağaçları

Karar ağaçları, sınıflandırma ve tahmin için sıkça kullanılan bir veri madenciliği yaklaşımıdır. Sinir ağları gibi diğer metodolojilerin de sınıflandırma için kullanılıyor olmasına rağmen karar ağaçları, kolay yorumu ve anlaşılabilirliği açısından karar vericiler için avantaj sağlamaktadır (Chien ve Chen, 2008: 281).

Karar ağaçlarının hedefi bağımlı değişkendeki farklılıkları maksimize edecek şekilde veriyi sıralı bir biçimde farklı gruplara ayırmaktır. Karar ağacı, adında belirtildiği şekilde ağaç görünümünde bir tekniktir. Karar düğümleri, dallar ve yapraklardan oluşmaktadır. Karar düğümünde, veriye uygulanacak test tanımlanır. Her düğüm bir özellikteki testi gösterir. Dal, testin sonucunu gösterir. Elde edilen her dal ile tanımlanacak sınıfın belirlenmesi amaçlanır. Ancak dalın sonucunda sınıflandırma tamamlanamıyorsa tekrar bir karar düğümü oluşur. Dalın sonucunda bir sınıflandırma elde edilebiliyorsa yaprak elde edilmiş olur. Yaprak, verileri kullanarak elde edilmek istenen sınıflandırmanın sınıflarından birini tanımlar (Argüden ve Erşahin, 2008: 47-48).

Karar ağacı tekniği ile verinin sınıflandırılması, öğrenme ve sınıflama olmak üzere iki basamaklı bir işlemde oluşmaktadır. Öğrenme basamağında önceden bilinen bir eğitim verisi, model oluşturmak amacıyla sınıflama algoritması tarafından analiz edilir. Öğrenilen model, sınıflama kuralları veya karar ağacı olarak gösterilir. Sınıflama basamağında ise test verisi, sınıflama kurallarının veya karar ağacının doğruluğunu belirlemek amacıyla kullanılır. Eğer doğruluk kabul edilebilir oranda ise, kurallar yeni verilerin sınıflanması amacıyla kullanılır (Özekes ve Çamurcu, 2002).

Karar ağacı tabanlı ilk algoritma ve yazılım, Morgan ve Sonquist adlı araştırmacılar tarafından 1970'li yılların başlarında önerilen AID algoritmasıdır (Emel ve Taşkın, 2005: 228). Temelleri AID (Automatic Interaction Detector) yöntemi ile atılan karar ağacı modelleri, ID3 (Iterative Dichotomiser 3; Quinlan, 1979), CHAID (ChiSquared Automatic Interaction Detection; Kass, 1980), CART (Classification Regression Trees; Breiman vd., 1984), C4.5 (Quinlan, 1993), SLIQ (Supervised Learning in Quest; Mehta vd., 1996), C5.0 (Quinlan, 1997), SPRINT (Scalable Parallelizable Induction of Decision Trees; Shafer vd., 1996) ve QUEST (Quick, Unbiased, Efficient Statistical Tree; Loh ve Shih, 1997) gibi çeşitli algoritmalar ile sürdürülmüştür.

Karar ağacı algoritmalarının kendilerine has özellikleri söz konusudur. Örneğin, CART algoritması ikili bölümlenmeyle, C5.0 algoritması ise çoklu bölümlenme ile ağaç oluşturmaktadır (Larose, 2014: 183). Yukarıda bahsetmiş olduğumuz algoritmaların özelliklerine kısaca değinecek olursak; CART, ağaç yapısına dayalı olarak sınıflandırma ve regresyon modellerinin türetilmesi için yaygın olarak kullanılan bir algoritmadır. CART ağaç modeli, tek değişkenli ikili kararların bir hiyerarşisini içermektedir. CART modellerinde bölünmelerin bulunmasında kategorik hedef değişkenler için Gini, sürekli hedef değişkenler için ise en küçük kareli sapma (LSD) kullanılabilir (Oğuzlar, 2004: 82-83). CART'da tanımlanan temel bölünme yöntemi C4.5'te de kullanılmaktadır. Ağaç yapısındaki farklılıklar, bölünme kriterlerinde, budama yönteminde ve kayıp değerlerin işleme biçiminde görülmektedir (Kantardzic, 2011:189). Tek değişkenli bölünmelere dayanan karar ağaçları üretmek için iyi bilinen algoritmalar arasında Quinlan'ın ID3 ve bu algoritmanın genişletilmiş versiyonu olan C4.5 algoritması yer almaktadır. ID3 ve C4.5 algoritmalarında öznitelik seçimi, bir düğümdeki örneklere uygulanan entropi ölçüsünün minimize edilmesine dayanmaktadır. Entropiye dayalı yaklaşım, örneklemin veritabanında sınıflandırılmasına izin verilecek test sayısını en aza indirmeyi gerektirmektedir (Kantardzic, 2011: 171-172). CHAID, öznitelik seçimi için Ki-kare testleri kullanan bir karar ağacı algoritmasıdır (Han vd., 2012: 343). SLIQ, hızlı ölçeklenebilir bir sınıflayıcıdır. Hızlı ağaç budama algoritması mevcuttur. SPRINT, büyük veri kümeleri için idealdir. Bölme işlemi tek bir niteliğin değerine dayanmaktadır. Tüm bellek sınırlamaları üzerinde nitelik listesi veri yapısı kullanarak işlem yapmaktadır (Emel ve Taşkın, 2005: 230).

Çalışmada en yüksek doğruluk oranı C5.0 algoritması ile elde edilmiş olup, bu algoritma ayrı bir başlık altında incelenmiştir.

3.1.1. C5.0 Algoritması

C5.0 algoritması ise C4.5'in geliştirilmiş hali olup, özellikle büyük veri setleri için kullanılmaktadır. C5.0 algoritması doğruluğu arttırmak için boosting algoritmasını kullandığından boosting ağaçları olarak da bilinir. C5.0 algoritması C4.5'e göre çok daha hızlı olup, hafızayı daha verimli kullanmaktadır. Her iki algoritmanın sonuçları aynı olsa da C5.0 biçim olarak daha düzgün karar ağaçları elde etmemizi sağlamaktadır.

C5.0 algoritması, tek bir düğümle başlamakta ve en uygun sınıflayıcının tespiti için bilgi kazanımı adı verilen buluşsal (heuristic) ve entropi tabanlı bir ölçü kullanmaktadır. Bu ölçüye göre belirlenen değişkenin her bir değeri dallara dönüşmekte ve devam eden süreçte kalan değişkenlerin dikkate alınması ile aynı işlem sürdürülmektedir. Bu süreç, belirli bir düğümdeki tüm örneklerin aynı sınıfa ait olması, bölümlenecek yeni bir değişkenin kalmaması ve ilgili düğüm için veri kümesinde en yüksek bilgi kazanımına, bir başka ifadeyle en büyük entropi azalımına sahip olan sınıma değişkeninin söz konusu değerine sahip kayıt bulunmaması durumlarında son bulmaktadır (Çakır, 2008: 61).

C5.0 algoritması, test edilecek özneteliği seçmek için Entropi kavramına dayalı "Bilgi Kazancı" kriterini kullanmaktadır. Bir eğitim kümesindeki sınıf niteliğinin alacağı değerler kümesi T ve n farklı değerle tanımlanan T_1, T_2, \dots, T_n alt veri kümeleri için olası bir sınıma değişkeninin (X) seçilmesi sorununu ele alalım. Herhangi bir örnek kümesi S olmak üzere, k adet farklı sınıf üzerinden C_i sınıfındaki örneklerin frekansı $f(C_i, S)$, S veri kümesine ait örneklerin sayısı ise $|S|$ ile gösterilmek üzere, veri kümesinin bölümlenmesi için gereken bilginin beklenen değeri yani entropi Eş. 1'deki gibi hesaplanmaktadır:

$$\text{Entropi}(T) = - \sum_{i=1}^k ((f(C_i, T) / |T|) \cdot \log_2(f(C_i, T) / |T|)) \quad (1)$$

Benzer şekilde X değişkeni ile gerçekleştirilen bölümlenme sonrasında T sınıf değerlerinin herbiri için entropi Eş. 2'deki gibi hesaplanmaktadır:

$$\text{Entropi}_x(T) = \sum_{i=1}^n ((|T_i| / |T|) \cdot \text{Entropi}(T_i)) \quad (2)$$

Sonuç olarak X değişkeni ile elde edilen bilgi kazancı Eş. 3'teki gibi hesaplanmaktadır:

$$\text{Kazanç}(X) = \text{Entropi}(T) - \text{Entropi}_x(T) \quad (3)$$

Kazanç kriterine göre en yüksek Bilgi Kazancı değerine sahip olan değişken sınıma değişkeni olarak seçilmektedir (Kantardzic, 2011: 175).

4. Bulgular

Uçak seferlerinde yaşanan rötarları etkileyen faktörlerin analiz edildiği bu çalışmada gecikmeler bağımlı değişkeni ifade ederken, gelen-giden uçaklar, pist numaraları, uçakların kategorisi, saat dilimi, gün ve hat ise bağımsız değişkenleri ifade etmektedir. Değişkenlere ait açıklamalar Tablo 1'de verilmektedir.

Atatürk Havalimanı'nda; 17-35 doğrultusunda birbirine paralel iki adet ve 05-23 doğrultusunda bu paralel pistlerle V şeklini oluşturan bir adet pist ile birlikte toplam üç adet pist bulunmaktadır. Atatürk Havalimanı'nda aktif pist seçimi iki ölçüt dikkate alınarak yapılmaktadır. Birinci ölçüt rüzgâr yön kriteri, ikinci ölçüt rüzgâr şiddet kriteridir. Rüzgâr şiddetinin güney bileşeni 4 Knot'dan küçük ise aktif pist olarak kalkışlar 35 doğrultusunda, inişler 05 doğrultusunda gerçekleştirilmektedir.

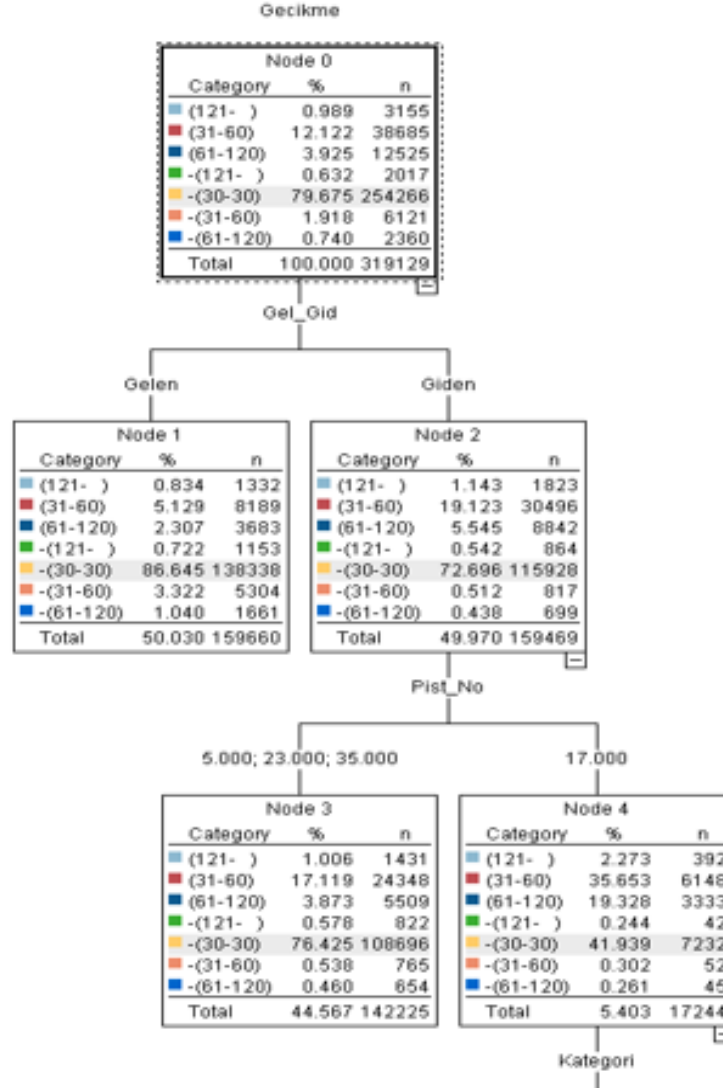
Rüzgâr şiddetinin güney bileşeni 4 Knot'dan büyük ise pist olarak kalkışlar 17 doğrultusunda inişler 23 doğrultusunda gerçekleştirilmektedir. Bu durumda bir uçak iniş operasyonunu tamamlamadan kalkış operasyonuna, kalkış operasyonunu tamamlamadan da iniş operasyonuna izin verilmemektedir. Operasyonlardan 30 dakika önce ve 30 dakika sonra süreler içinde gerçekleştirilen operasyonlar zamanında gerçekleştirilmiş sayılmaktadır. (Horonjeff, Sproule ve Young, 2010).

Tablo 1: Bağımlı Değişken ve Bağımsız Değişkenler

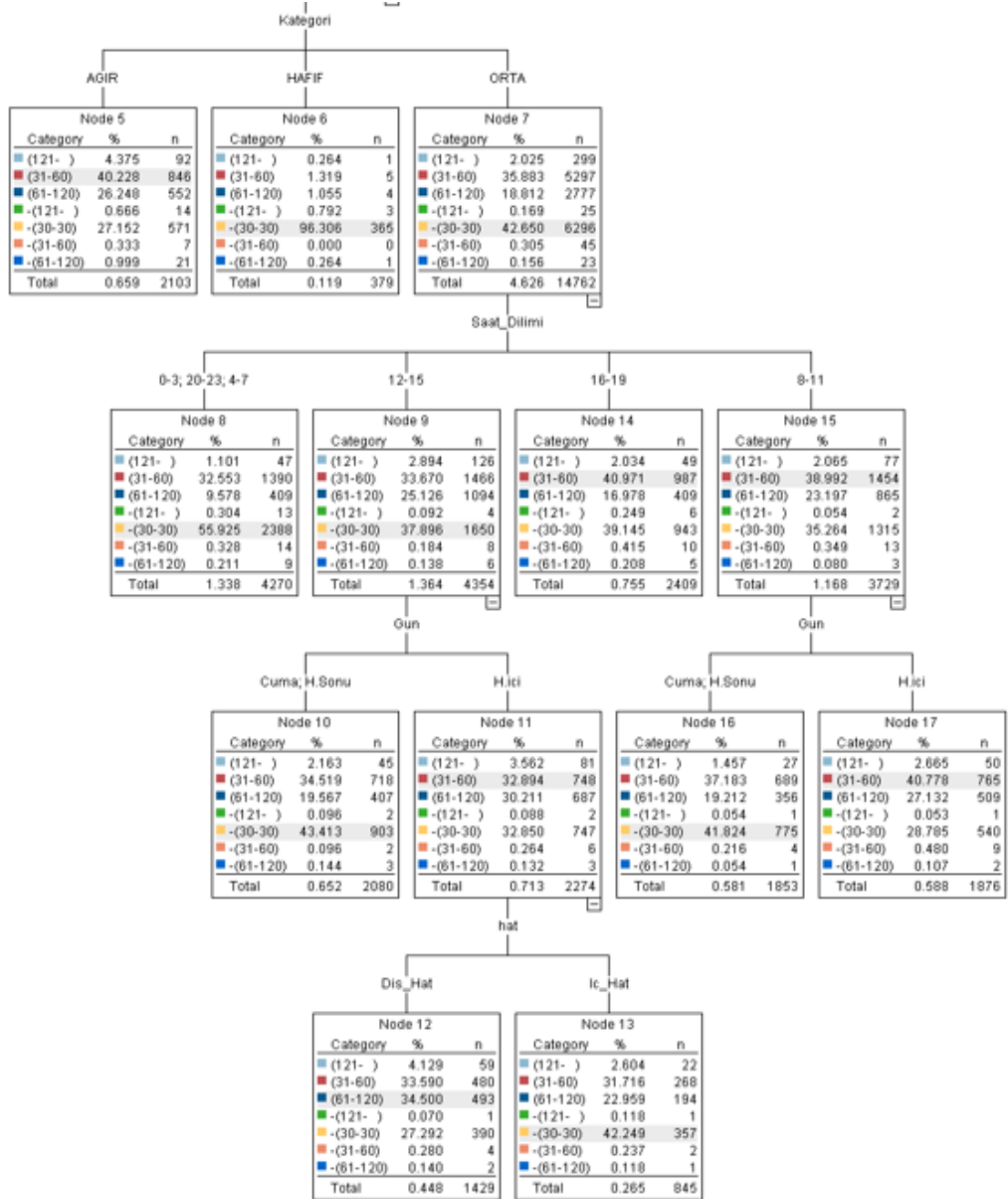
DEĞİŞKEN	AÇIKLAMA
Gecikme	
-(121-)	121 dk.'dan fazla erken gelme durumu bu kategoride tanımlanmıştır.
-(61-120)	61-120 dakika arası erken gelme durumu bu kategoride tanımlanmıştır.
-(31-60)	31-60 dakika arası erken gelme durumu bu kategoride tanımlanmıştır.
-(30-30)	Zamanında gerçekleştirilen iniş ve kalkışlar bu kategoride tanımlanmıştır.
(31-60)	31-60 dakika arası geç gelme durumu bu kategoride tanımlanmıştır.
(61-120)	61-120 dakika arası geç gelme durumu bu kategoride tanımlanmıştır.
(121-)	121 dk.'dan fazla geç gelme durumu bu kategoride tanımlanmıştır.
Gel-Gid	
Gelen	Atatürk Havalimanına gelen uçaklar bu kategoride tanımlanmıştır.
Giden	Atatürk Havalimanından giden uçaklar bu kategoride tanımlanmıştır.
Pist No	
05	Atatürk Havalimanı'nda; 17-35 doğrultusunda birbirine paralel iki adet ve 05-23 doğrultusunda bu paralel pistlerle V şeklini oluşturan bir adet pist ile birlikte toplam üç adet pist bulunmaktadır. 05 Pist başının derecesini göstermektedir.(50 derece)
17	Atatürk Havalimanı'nda; 17-35 doğrultusunda birbirine paralel iki adet ve 05-23 doğrultusunda bu paralel pistlerle V şeklini oluşturan bir adet pist ile birlikte toplam üç adet pist bulunmaktadır. 17 Pist başının derecesini göstermektedir.(170 derece)
23	Atatürk Havalimanı'nda; 17-35 doğrultusunda birbirine paralel iki adet ve 05-23 doğrultusunda bu paralel pistlerle V şeklini oluşturan bir adet pist ile birlikte toplam üç adet pist bulunmaktadır. 23 Pist başının derecesini göstermektedir.(230 derece)
35	Atatürk Havalimanı'nda; 17-35 doğrultusunda birbirine paralel iki adet ve 05-23 doğrultusunda bu paralel pistlerle V şeklini oluşturan bir adet pist ile birlikte toplam üç adet pist bulunmaktadır. 35 Pist başının derecesini göstermektedir.(350 derece)
Kategori	
Ağır	Azami kalkış ağırlığı 136.000 kilogram (300.000 pound) veya daha ağır olan uçaklar bu kategoride tanımlanmıştır.
Orta	Azami kalkış ağırlığı 7.000 kilogram (15.000 pound)'dan ağır ve 136.000 kilogram (300.000 pound)'dan hafif olan uçaklar bu kategoride tanımlanmıştır.
Hafif	Azami kalkış ağırlığı 7.000 kilogram (15.000 pound) veya daha hafif olan uçaklar bu kategoride tanımlanmıştır.
Saat Dilimi	
0-3	Saat 00:00 ile 03:59 arası saat dilimini ifade etmektedir.
4-7	Saat 04:00 ile 07:59 arası saat dilimini ifade etmektedir.
8-11	Saat 08:00 ile 11:59 arası saat dilimini ifade etmektedir.
12-15	Saat 12:00 ile 15:59 arası saat dilimini ifade etmektedir.
16-19	Saat 16:00 ile 19:59 arası saat dilimini ifade etmektedir.
20-23	Saat 20:00 ile 23:59 arası saat dilimini ifade etmektedir.
Gün	
H.İç	Hafta içi günleri ifade etmektedir.
Cuma	Cuma gününü ifade etmektedir.
H.sonu	Cumartesi ve Pazar günlerini ifade etmektedir.
Hat	
Dış Hat	Yurtdışı yapılan uçuşları ifade etmektedir.
İç Hat	Yurtiçi yapılan uçuşları ifade etmektedir.

Çalışma kapsamında Access veri tabanından SQL veritabanına taşınan 319.129 veri SPSS Clementine 12.0 programı kullanılarak analiz edilmiş ve problem için en yüksek doğruluk oranını veren C5.0 algoritması ile kural çıkarımı Şekil 1'deki gibi gerçekleştirilmiştir.

Şekil 1: C5.0 Algoritması ile Elde Edilen Karar Ağacı



Şekil 1 (Devam): C5.0 Algoritması ile Elde Edilen Karar Ağacı



Toplam 319.129 verinin analiz edildiği çalışmada; Havalimanından giden uçaklar arasında 05, 23 ve 35 yönünde kalkış gerçekleştiren trafiklerin %76' sının; 17 yönünde kalkış gerçekleştiren trafiklerin ise %42'sinin 30 dk. erken veya 30 dk. geç süreler içerisinde kalkış faaliyetlerini gerçekleştirdiği görülmektedir. Bu nedenle uzmanlarının da görüşü alınarak 17-35 pistine paralel bir pist yapılması veya Atatürk Havalimanı için 4 Knot kabul edilen sakin rüzgâr limitinin 8 Knot'a çıkartılması önerilmektedir. Ayrıca Hafif kategorideki uçakların %96'sının, Orta kategorideki uçakların %43'ünün, Ağır kategorideki uçakların ise sadece %27'sinin 30 dakikalık süreler içerisinde kalkış faaliyetlerini gerçekleştirdiği görülmektedir. Bu doğrultuda alan uzmanlarının da görüşü alınarak Güney yönlü rüzgâr altında "23 istikametinde iniş yapacak uçak var ise 17 pistinden

kalkışlara izin verilmez” kuralının; “23 istikametinde iniş yapacak sadece büyük kategoride uçak var ise 17 pistinden kalkışlara izin verilmez” olarak değiştirilmesi önerilmektedir.

5. Sonuç

Havalimanı kapasitesinin sınırlı olması ve hava trafik yoğunluğu havayolu şirketlerinin maliyetlerinde artışa sebep olurken, müşteri memnuniyetinde de azalmaya yol açmaktadır. Havalimanı kapasitesini kısa ve orta vadede arttırmanın yolu, uluslararası havacılık kurallarına uygun olarak hava trafik kurallarında yapılacak değişikliklerdir. Bu çalışmada güney yönlü rüzgârlar (Iodos) altında kapasitesinde azalma meydana gelen Atatürk Havalimanı'nın kapasitesini arttırmak amacıyla uçak seferlerinde yaşanan gecikmeleri etkileyen faktörlerin analizi için veri madenciliği uygulamasına yer verilmiştir.

Veri madenciliği çalışması sonucunda uçak seferlerinde yaşanan gecikmeleri en çok uçağın geliş-gidiş durumlarının etkilediği görülmüştür. Özellikle giden trafik durumunun gecikmeleri etkilediği görülmektedir. Giden trafiklerden, güney yönlü rüzgarlar altında (17 pistinin kullanıldığı durumlarda) gecikmelerin arttığı tespit edilmiştir. Bu tür operasyonların gerçekleştiği durumlarda ise gecikmelerden en çok büyük ve orta kategorideki uçaklar etkilenmektedir. Ayrıca gecikmeler hafta içi ve yurt dışına gerçekleştirilen 08-11 ve 16-19 saatleri arasında gerçekleştirilen uçuşlarda görülmektedir.

Veri madenciliği ile tespit edilen sonuçlara dayanarak hava trafik faaliyetlerinin düzenlenmesi sonucunda veya simülasyon yardımıyla modellenmesi yapıldığı takdirde yaşanan gecikmelere ekonomik ve kısa zaman alan çözümler bulunabileceği öngörülmektedir.

Kaynakça

- Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 1, 1-22.
- Argüden, Y. ve Erşahin, B. (2008). *Veri Madenciliği, 1. Baskı*. ARGE Danışmanlık Yayınları, İstanbul.
- Chien, C.F. ve Chen, L.F. (2008). Data Mining to Improve Personnel Selection and Enhance Human Capital: A Case Study in High-Technology Industry. *Expert Systems with Applications*, 34, 280-290.
- Ching W. K. ve Pong M. K. (2002). *Advances in Data Mining and Modeling (1st ed.)*. World Scientific, Hong Kong.
- Çakır, Ö. (2008). Veri Madenciliğinde Sınıflandırma Yöntemlerinin Karşılaştırılması: Bankacılık Müşteri Veri Tabanı Üzerinde Bir Uygulama, Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Dolgun M. Ö. ve Ersel D. (2014). Doğrudan Pazarlama Stratejilerinin Belirlenmesinde Veri Madenciliği Yöntemlerinin Kullanımı, *İstatistikçiler Dergisi: İstatistik&Aktüerya*, 7, 1-13.
- Duran A. E., Pamukcu A. ve Bozkurt H. (2014). “Comparison of Data Mining Techniques for Direct Marketing Campaigns”, *Sigma Journal of Engineering and Natural Sciences*, 32, 142-152.
- Emel, G. G. ve Taşkın, Ç. (2005). Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6(2), 221-239.
- Fan, H., Tarun, P., Shih, D., Kim, S., Chen, V., Rosenberger, J. ve Bergman, D. (2011). Data Mining Modeling on the Environmental Impact of Airport Deicing Activities. *Expert Systems with Applications*, 38, 14899-14906.
- Han J., Kamber M. (2006). *Data Mining: Concepts and Techniques (2nd ed.)*. Morgan Kaufmann, USA.

- Han, J., Kamber, M. ve Pei, J. (2012). *Data Mining: Concepts and Techniques (3rd ed.)*. Morgan Kaufmann, USA.
- Horonjeff, R., F., M., Sproule, W. ve Young, S. (2010). *Planning and Design of Airports* (s. 483-510). California, Berkeley: McGraw-Hill Company.
- Jacobs P. (1999). *Data Mining: What General Managers Need to Know*. Harvard Management Update, 4(10), 8-9.
- Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, and Algorithms (2nd ed.)*. Wiley, Canada.
- Köktürk, F., Ankaralı, H. ve Sümbüloğlu, V. (2009). Veri Madenciliği Yöntemlerine Genel Bakış. *Türkiye Klinikleri J Biostat*. 1. 20-25.
- Larose, D. T. ve Larose, C. D. (2014). *Discovering Knowledge in Data (2nd ed.)*. Wiley, Canada.
- Linoff G.S. ve Berry M.J.A. (2011). *Data Mining Techniques for Marketing, Sales and Customer Relationship Management (3rd ed.)*. Wiley, Canada.
- Liou, J., Tang, C., Yeh, W. ve Tsai, C. (2011). A decision Rules Approach for Improvement of Airport Service Quality. *Expert Systems with Applications*, 38, 13723-13730.
- Oğuzlar, A. (2004). CART Analizi ile Hanehalkı İşgücü Anketi Sonuçlarının Özetlenmesi. *Atatürk Üniversitesi İİBF Dergisi*, 18, 79-90.
- Özçakır, F. C. ve Çamurcu, A. Y. (2007). Birliktelik Kuralı Yöntemi için Bir Veri Madenciliği Yazılımı Tasarımı ve Uygulaması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 12, 21-37.
- Özekes, S. ve Çamurcu, A.Y. (2002). Veri Madenciliğinde Sınıflama ve Kestirim Uygulaması, *Marmara Üniversitesi Fen Bilimleri Dergisi*, 18, 1-17.
- Savaş, S., Topaloğlu., N. ve Yılmaz, M. (2012). Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 21, 1-23.
- Siknun, G. P. ve Sitanggang, I. S. (2016). Web-Based Classification Application for Forest Fire Data Using the Shiny Framework and the C5.0 Algorithm, *Procedia Environmental Sciences*, 33, 332-339.
- Smith, L.D. ve Ehmke, J. F. (2016). A Mathematical Programming Technique for Matching Time-Stamped Records in Logistics and Transportation Systems. *Transportation Research Part C*, 69, 375-385.
- Suner, A. ve Çelikoğlu, C. C. (2010). Toplum Tabanlı Bir Çalışmada Çoklu Uygunluk Analizi ve Kümeleme Analizi ile Sağlık Kurumu Seçimi, *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 25, 43-55.
- Taşkın Ç. ve Emel G. G. (2010). Veri Madenciliğinde Kümeleme Yaklaşımları ve Khonen Ağları ile Perakendecilik Sektöründe Bir Uygulama, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 15(3), 395-409.
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers.

ANALYSIS OF THE FACTORS AFFECTING THE FLIGHT DELAYS

Extended Abstract

Aim: The air traffic volume, which has been increasing in the whole world, is being met by airports with limited resource capacity. The limited capacity of the airport and the increase in the volume of air traffic volume are leading to delays at the airport. These delays cause an increase in the costs of airline companies and also cause a decrease in customer satisfaction. The way to increase airport capacity in the long run is to build a new runway or high-capacity airport. Another way to increase airport capacity in the short and medium lane is to make changes to the air traffic rules in accordance with international aviation regulations. In this study, it is aimed to analyze the factors affecting the delays experienced in aircraft flights in order to increase the airport capacity which is decreased in the capacity under the southwest.

Method: In this study, decision tree which is one of the classification methods of data mining, is used and the factors affecting the delays in aircraft flights are analyzed. Classification is the area where data mining is most commonly used. The aim is to predict which class of non-classifiable data will be included in the class, based on class-set data.

The decision tree is a data mining approach that is frequently used in classification and estimation. Despite being capable of being used in classification of other methodologies like the nerve networks, the decision trees with their easy to make interpretations and ease of being understood provides advantage or decision makers. The most widely used decision tree algorithm is the C4.5 algorithm which is the develop state of ID3 algorithm that was proposed in 1986 by Quinlan. The C5.0 algorithm is the develop state of C4.5 and it is used especially for large data sets. To increase the correctness for the C5.0 algorithm, the boosting algorithm is used and therefore they are also known as boosting trees. The C5.0 algorithm is more rapid as compared to C4.5 and uses memory in a more productive manner. Even the results of both of the two algorithms are the same, the C5.0 as form makes it possible to come out with a smoother decision tree.

Findings: In this study in which a total of 319,129 data were analyzed; among the aircraft departing from the airport, 76% of the traffic departing in 05, 23 and 35 direction; of the traffic that departs in the 17 direction, 42% of the traffic takes off 30 minutes early or 30 minutes later. For this reason, it is suggested to make a runway parallel to the 17-35 runway by taking the opinion of the experts or to remove the minimum wind limit which is accepted as 4 Knot for Atatürk Airport to 8 Knot. It is also seen that 96% of the aircraft in the light category, 43% of the aircraft in the middle category, and only 27% of the aircraft in the heavy category, take off activities within 30 minutes. In this direction, taking the opinion of experts in the area, "the departure from 17 runway is not allowed if there is a plane landing in 23 direction" under the south direction wind; if there are only large categories of airplanes to land in 23, it is recommended that" departures from 17 runways are not allowed".

Conclusion: As a result of the data mining study, it is seen that the delays experienced by airplane flights are most affected by the arrival/departure status of the aircraft, runway numbers, and categorizations. The result is that the effects of the time slice, day and line variables on delays are not significant. It is evaluated that economical and short time-consuming solutions can be found when solving the constraints determined by the data mining with simulation help.