

Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması

Comparison of Apriori and FP-Growth Algorithms on Determination of Association Rules in Authorized Automobile Service Centres

Yrd. Doç. Dr. Semra Erpolat

Öz

Veri Madenciliği, bir veri kümesi içerisinde keşfedilmiş örüntüleri bulmayı hedefleyen teknikler bütünüdür. Veri Madenciliği'nin amacı, geçmiş faaliyetlerin analizini temel alarak gelecekteki davranışların tahminine yönelik karar verme modelleri yaratmaktır.

Bu çalışmada Türkiye'de otomotiv sektöründe faaliyet gösteren bir yetkili servisin müşterilerine ait alış-veriş verileri, Apriori ve FP-Growth Algoritmaları kullanılarak analiz edilmiştir. Böylelikle müşterilerin hangi ürünleri birlikte satın aldıkları gözlemlenmiş ve bu doğrultuda karı artırmaya yönelik uygulanacak kampanya va promosyonlara yön verilmeye çalışılmıştır.

Anahtar Kelimeler: Veri Madenciliği, Birliktelik Kuralları, Apriori Algoritması, FP-Growth Algoritması, Pazar Sepet Analizi

Abstract

Data Mining is used to describe the totality of techniques which aim to find the unexplored patterns in a set of data. The purpose of data mining is to create models of decision-making devoted to estimations of future behavior based on analysis of past activities.

In this study the shopping data of the customers of an authorized service, operating in the automotive sector in Turkey, were analyzed using Apriori and FP-Growth Algorithms. This way, it is observed which products were purchased together by customers and in line with this observation, campaigns and promotions were given a direction to increase the profit.

Keywords: Data Mining, Association Rules, Apriori Algorithm, FP-Growth Algorithm, Market Basket Analysis

Giriş

Dijital verilerin her geçen gün daha da artması ve büyük veri tabanlarında saklanması, zamanla bu verilerden en verimli şekilde faydalanma ihtiyacını doğurmuştur. Bu ihtiyaç, "Veri Tabanlarında Bilgi Keşfi (VTBK)" adı altında sürekli ve yeni arayışlar ortaya çıkarmıştır. VTBK bir süreç olup farklı kısımlardan oluşmaktadır. Sürecin model kurma ve değerlendirme aşamalarını oluşturan en önemli kısmı ise "Veri Madenciliği (VM)"dir.

VM, büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanılarak aranmasıdır (Alpaydın, 2000, s.4).

Kural çıkarma, VM'nin en önemli işlevlerinden bir tanesidir. Geleceğin, en azından yakın geleceğin, geçmişten çok fazla farklı olmayacağını varsayarsak geçmiş veriden çıkarılmış olan kurallar gelecekte de geçerli olacak ve ilerisi için doğru tahmin yapmamızı sağlayacaktır (Alpaydın, 2000, s.7). Kural çıkarmak amacıyla farklı yöntemler geliştirilmiştir. Bu yöntemlerden en bilinenleri Apriori ve FP-Growth Algoritmalarıdır.

Literatürde Apriori ve FP-Growth Algoritmaları kullanılarak çeşitli çalışmalar gerçekleştirilmiştir. Örneğin; Vishal S., Nikita J., Sharad V. (2010) Apriori Algoritması'nı kullanarak bir süper markete ait veri tabanında yer alan bilgiler ışığında satılan ürünlere ilişkin birliktelik kuralları çıkarmışlardır. Surendiran R., Rajan.K.P. ve Sathish K.M. (2010), FP-Growth Al

goritması ile mobil servis kullanıcıları için birliktelik kuralları oluşturmuşlardır. Wei Z., Hongzhi L. ve Na Z. (2008) Apriori Algoritması ile çözümlenmiş bilinen iki problemi FP-Growth Algoritması'nı kullanarak sonuca ulaştırmışlardır. Elde ettikleri bulgulardan FP-Growth Algoritması'nın Apriori Algoritması'na göre daha etkin ve tutarlı sonuçlar verdiğini ayrıca zaman açısından da daha hızlı olduğunu gözlemişlerdir. Daniel H. (2011) ise Apriori ve FP-Growth Algoritmaları'nın birliktelik kuralları çıkarmadaki performanslarını karşılaştırmak amacıyla internet üzerinden alış-veriş yapılan bir mağazaya ait verileri kullanmış ve elde ettiği bulgulardan FP-Growth Algoritması'nın daha iyi sonuçlar verdiğini vurgulamıştır.

Apriori ve FP-Growth Algoritma'larının karşılaştırıldığı bu çalışmada, otomotiv sektöründe faaliyet gösteren bir yetkili servisin müşterilerine ait alış-veriş verileri kullanılarak servisin müşteri eğilimleri belirlenmeye çalışılmıştır. Elde edilen bulgulardan servis yetkililerine müşterileri ile ilgili bilgiler verilmeye çalışılarak satışları artırmaya yönelik kararlar alınmalarında destek sağlanmıştır.

Veri Madenciliği

VM, eldeki verilerden üstü kapalı, çok net olmayan, önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin çıkarılmasını sağlar. Bir başka ifadeyle büyük miktardaki verinin analiz edilerek anlamlı şablon ve kuralların keşfedilmesine imkân verir (Berry ve Linoff, 2004, s.12). Pek çok alanda yaygın olarak kullanılan VM, Türkiye'de ve dünyada özellikle son yıllarda oldukça önem kazanmış bir tekniktir.

VM, temelde iki ana başlıkta incelenmektedir. Birincisi, elde edilen örüntülerden sonuçları bilinmeyen verilerin tahmini için kullanılan tahmin edici, diğeri ise eldeki verinin tanımlanmasını sağlayan tanımlayıcıdır (Akpınar, 1998, s.5). VM'de kullanılan modeller;

kümeleme analizi (tanımlayıcı), birliktelik kuralları ve ardışık zamanlı örüntüler (tanımlayıcı), sınıflandırma (tahmin edici) ve regresyon analizi ve zaman serileri analizi (tahmin edici) olarak belirtilebilir.

Birliktelik Kuralları

Veri tabanlarından birliktelik kurallarının bulunması VM'nin en önemli konularından biri olup, bir arada sık olarak görülen ilişkilerin ortaya çıkarılmasını ve özetlenmesini sağlar. Örneğin, bir alış-veriş sırasında müşterinin hangi ürün veya hizmetleri satın almaya eğilimli olduğunun belirlenmesi, müşteriye daha fazla ürünün satılmasını sağlayarak şirket kârını arttırıcı rol oynar.

Satın alma eğilimlerinin tanımlanmasını sağlayan birliktelik kuralları ve ardışık zamanlı örüntüler, pazarlama amaçlı olarak Pazar Sepet Analizi adı altında VM'de sıkça kullanılmaktadır. Pazar Sepet Analizi, müşterilerin alış-veriş alışkanlıklarının veri tabanındaki bilgiler aracılığıyla ortaya çıkartılması işlemidir. (Silahtaroglu, 2008, s.22).

Birliktelik kurallarının en önemli uygulama alanları arasında; Pazar Sepet Analizi, çapraz-pazarlama, promosyon analizleri, katalog ve yerleşim düzeni tasarımları bulunmaktadır.

Birliktelik Kuralları İçin Kullanılan Algoritmalar

Büyük veri yığınlarından bilgi sağlamaya yarayan birçok algoritma ardışık veya paralel olarak sınıflandırılabilir. Ardışık algoritmalar ürün kümelerinin oluşturulduğu ve sayıldığı mantıksal ifadeler içerir. Diğer taraftan, paralel algoritmalar ise büyük ürün kümelerinin paralellik sağlayarak oluşturulmasını sağlar. Tablo 1'de ardışık algoritmalar ile paralel ve dağıtılmış algoritmalar sıralanmıştır. Çalışmada ardışık algoritmalarından Apriori ve FP-Growth Algoritmaları incelenmiştir. İzleyen alt kesimlerde söz konusu algoritmaların ayrıntılarına değinilecektir.

Tablo 1. Birliktelik Algoritmaları

Ardışık Algoritmalar	Paralel ve Dağıtılmış Algoritmalar
AIS	CD (Sayım Dağılımı)
SETM	PDM (Paralel VM)
Apriori	CCPD (Ortak Aday Bölünmüş Veri Tabanı)
Apriori-TID	DD (Veri Dağılımı)
Apriori-Hybrid	IDD (Akıllı Veri Dağılımı)
OCD (Sıraışı Aday Belirleme)	HPA (Bağlantı Kurallarının Çırpı Temelli Paralel Madenciligi)
Bölümleme Tekniđi	PAR (Paralel Bağlantı Kuralları)
Örnekleme Tekniđi	DMA (Dağıtılmış Madencilik Algoritması)
CARMA (Sürekli Bağlantı Kuralı Madenciligi)	Candidate Distribution (Aday Dağılımı)
DIC(Dinamik Nesne Sayımı)	SH (Çarpık Taşıma)
FP-Growth	HD (Hibrid Dağılımı)

Apriori Algoritması

Apriori Algoritması, bilgileri bir önceki adımdan aldığı için “prior (önceki)” kelimesinden türetilmiştir. Birliktelik kuralı çıkarım algoritmaları içerisinde en fazla bilinen algoritma olan Apriori Algoritması geniş nesne kümelerinin ortaya çıkartılması işlemleri için kullanılır (Agrawal, vd., 1993, s.490). Bir nesne kümesinin geniş olarak adlandırılabilmesi için ise o nesne kümesinin kullanıcı tarafından verilen minimum destek seviyesinin üzerinde bir destek seviyesine sahip olması gerekir. Geniş nesne kümelerini ortaya çıkartan algoritmalar eldeki tüm verileri birçok kez tararlar. İlk taramada, her bir nesnenin destek seviyesi hesaplanarak kullanıcı tarafından başlangıçta girilen minimum destek seviyesi ile karşılaştırılır ve her bir nesnenin geniş olup olmadığına bakılır. Bundan sonraki her tarama bir önceki taramada geniş olarak tespit edilmiş nesnelere başlar ve geniş nesne kümeleri oluşturulur. Bu geniş nesne kümelerine aday nesne kümeleri denir. Taramanın sonunda ise hangi aday nesne kümesinin gerçekten geniş olduğu kontrol edilir. Bir sonraki taramada, yine bir önceki taramada geniş olarak seçilen nesne kümelerinden başlanır ve veri tabanının sonuna kadar bu nesne kümelerin destekleri hesaplanır. Bu işlem başka yeni geniş nesne kümeleri bulunamayana kadar sürer.(Agrawal ve Srikant, 1994, s.495)

Pazar Sepet Analizi çözümlerinde satılan ürünler arasındaki ilişkileri ortaya koymak için “destek” ve “güven” gibi iki ölçütten yararlanılır. ‘Kural destek ölçütü’ bir ilişkinin tüm alış-verişler içinde hangi oranda tekrarlandığını belirler. ‘Kural güven ölçütü’, A ürün grubunu alan müşterilerin B ürün grubunu da alma olasılığını ortaya koyar. A ürün grubunu alanla

rın B ürün grubunu da alma durumu, yani birliktelik kuralı $A \rightarrow B$ biçiminde gösterilir. Bu durumda kural destek ölçütü şu şekilde ifade edilebilir:

$$\text{Destek}(A \rightarrow B) = \text{sayı}(A, B) / N \quad (1)$$

Burada $\text{sayı}(A, B)$ destek sayısı A ve B ürün gruplarını birlikte içeren alış-veriş sayısını göstermektedir. N ise tüm alış-verişlerin sayısını göstermektedir. A ve B ürün gruplarının birlikte satın alınması olasılığını ifade eden kural güven ölçütü aşağıdaki şekilde hesaplanır (Özkan, 2008, s.157):

$$\text{Güven}(A \rightarrow B) = \text{sayı}(A, B) / \text{sayı}(A) \quad (2)$$

Elde olan veride ürünler için sadece alındı/alınmadı bilgisi varsa sepet analizinde ürünler arasındaki bağlantı, destek ve güven kriterleri aracılığı ile hesaplanır. İki ürünün satın alınmasındaki bağıntının önemli olması için her iki değerin de olabildiğince yüksek olması gerekmektedir.

FP-Growth Algoritması

Birliktelik kuralları analizini gerçekleştirmek üzere geliştirilmiş yöntemlerden bir tanesi de FP-Growth (Frequent Pattern Growth) Algoritması’dır (Han ve Kamber, 2000, s.150). Diğer algoritmalara göre daha yüksek performans gösteren algoritma (Györödi, C., vd., 2004, s.220) tüm veri tabanını FP-Tree (Frequent Pattern Tree) adı verilen sıkıştırılmış bir ağaç veri yapısında tutar ve veri tabanını sadece iki kez tarar. Birinci taramada, tüm nesnelere destek değerlerini hesaplar, ikinci taramada ise ağaç veri yapısını oluşturur. Algoritmayı diğer algoritmalarından ayıran en

önemli özellikleri yaygın nesne kümelerini aday nesne kümeleri üretmeden test edebilmesi, büyük veri kümelerinde hızlı çalışabilmesi ve sistem kaynaklarını verimli kullanabilmesidir. FP-Algoritması'nda öncelikle veri tabanındaki her bir nesnenin destek değerleri hesaplanır. Destek değerleri, algoritmaya girdi olarak verilen destek eşik değerinden büyük ve eşit olan nesnelere büyükten küçüğe sıralanarak bir liste içerisine konulur. Böylelikle yaygın olmayan nesnelere ağaca eklenmesi önlenmiş olur. Daha sonra, veri tabanındaki her bir hareket kaydı nesnelere destek değerlerine göre sıralanarak ağaca sıkıştırılmış biçimde eklenir. Sıralama işlemi sayesinde destek değeri daha büyük olan nesnelere köke daha yakın olur. Sıkıştırma işlemi çok tekrarlı nesnelere ilk ekler olarak birleştirilmesi ile gerçekleştirilir. Bu metod arama maliyetini önemli ölçüde azaltır. Hareket kaydı içerisinde yer alan bir nesne ağaçta yoksa o nesne için yeni bir düğüm oluşturulur ve destek değeri 1 olarak atanır. Eğer o nesne daha önce ağaçta oluşturulmuş ise sadece o düğümün destek değeri 1 artırılır. Nesnelere ağaçtaki başlangıç noktaları "başlık tablosu" içerisinde tutulur.

Ağaç oluşturulduktan sonra üzerinde Growth Algoritması çalıştırılır. Growth Algoritması'nda öncelikle ele alınan nesnenin içerisinden geçtiği dallar belirlenir. Eğer tek bir dal varsa yaygın nesnelere kümesi, dalı oluşturan nesnelere kombinasyonudur. Eğer birden fazla dal varsa, destek değeri o daldaki minimum destek değeri olarak belirlenir. Daha sonra bu dallar o nesne için koşullu örüntü temelini oluşturur. Her bir koşullu örüntü temelinden koşullu örüntü ağacı elde edilir. Daha sonra bu şartlı örüntü ağacı üzerinden algoritma özyinelemeli olarak yeniden çalıştırılır (Han, J., vd., 2000, s.45). Birlikte sıklıkla görünen nesnelere kümesini belirleyen FP-Growth Algoritması, "böl ve yönet yaklaşımına" uygun olarak büyük yaygın nesnenin kendi içinde daha küçük nesne kümelerine ayrılmasını sağlar. Bu nedenle oluşturulan FP-Tree veri yapısı asıl veri kümesinden daha büyük olamaz (Pandey, A., vd., 2009, s.160).

Apriori Algoritması ile FP-Growth Algoritması'nın Karşılaştırılması

Aday nesne üretimli bir yöntem olan Apriori Algoritması, veri tabanındaki hareketleri göz önünde bulundurmadan bir önceki geçişte yaygın olarak belirlenmiş nesne kümelerini kullanarak, bir sonraki geçişte sayılacak olan aday nesne kümelerini oluşturur. k tane nesneden oluşan aday nesne kümeleri ($k-1$) tane nesne içeren yaygın nesne kümelerinin birleşiminden elde edilmektedir. Yaygın bir nesne kümesinin alt kümelerinin de yaygın olması gerekliliği mantığına dayanan Apriori Algoritması'nda işlemler, herhangi bir aday nesne kümesi elde edilemediği sürece devam eder.

FP-Growth Algoritması ise, yaygın nesne kümelerini aday nesnelere üretmeden bulmak için geliştirilmiş bir yöntemdir. Yöntemde, FP-Tree olarak adlandırılan ve veri tabanı tarama sayısını azaltmak üzere yaygın nesnelere oluşan bir ağaç yapısı kullanılmaktadır. Bu ağaç yapısında nesne kümelerinin birliktelik bilgileri yer alır. FP-Growth Algoritması, maliyeti yüksek olan aday nesne kümelerinin üretimini engellemek suretiyle arama maliyetini önemli ölçüde azaltır. Yapılan çalışmalar sonucunda büyük yaygın nesne kümelerini belirlemede FP-Growth Algoritması'nın Apriori Algoritması'na göre daha etkili, ölçeklenebilir bir yapıya sahip olduğu ve daha hızlı çalıştığı sonucuna varılmıştır. Ayrıca, ele alınan veri tabanının büyük, minimum destek sayısının ise küçük olduğu durumlarda FP-Growth Algoritması'nın Apriori Algoritması'na göre daha iyi sonuçlar verdiği görülmüştür. Şöyle ki; minimum destek sayısının küçük olması birçok nesnenin destek sayısını sağlamasına bu ise Apriori'de üretilen aday nesne küme sayısının artmasına neden olurken söz konusu artış daha etkili bir yapı olan örüntü madenciliğini kullanan FP-Growth için geçerli olmamaktadır.

Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesi

Bu çalışmada, Türkiye'de otomotiv sektöründe faaliyet gösteren yetkili servislerden bir tanesinin 2009 yılı Aralık ayı içinde yapmış olduğu otomobil parçaları satışları ele alınmıştır. Söz konusu servisin satı-

şını gerçekleştirebileceği toplam ürün sayısı 50'dir. Bu ürünler motor yağı, sürgülü cam mandalı, fren balatası, saplama bijon anahtar, somun bijon anahtar, krank mili, yağ keçesi, sprey temizleyici, antifriz, şanzıman, marş motoru, debriyaj çatalı, debriyaj baskı balata, debriyaj rulmanı, hava filtresi, karter yağı, boşaltma tapası, jant kapağı, termostat gövdesi, termostat su çıkış borusu, termostat yuvası, hidrolik fren yağı, fren vakum pompası, istavroz kutusu, defransiyel mahrutisi, hidrolik direksiyon yağı, rulman, ayna mahrutisi, şaft askısı, vites kolu, debriyaj merkez silindiri, denge çubuk burçları, geri vites lambası, yağ filtresi, kapı makarası, fren balata temizleyicisi, direksiyon pompası, silecek kolu, çamurluk, deflektör, tampon başlığı, direksiyon kolunu, amortisör, kaporta, stop lambası, viraj demiri, tekerlek, direksiyon dişlisi, yakıt pompa valfi, egzost manifoldu, aynadır. Belirlenen zaman diliminde firmanın satışları sonucunda toplam 40 fatura hazırlanmıştır. Bu faturalardan elde edilen bilgiler çalışmanın veri kümesini oluşturmuştur. Veri kümesi, her bir ürünün ele alınan faturada yer alıp almamasına, başka bir ifade ile satışının gerçekleştirilip gerçekleştirilmemesine bağlı olarak oluşturulmuştur. Böylece, faturada yer alan ürün için "1"; faturada yer almayan ürün için ise "0" değeri verilmiştir. Daha sonra veri kümesini oluşturan ürün-

ler arasındaki birliktelik kurallarının belirlenmesinde Pazar Sepet Analizi ile "Apriori" ve "FP-Growth" Algoritmaları kullanılmıştır. Algoritmalar Weka ve Excel programları kullanılarak gerçekleştirilmiştir. Çalışmayı yapmadaki amaç veri topluluğu içerisinde birbiri ile ilgili olan ürün kümelerini bularak, satış işleminde bu ürünlere uygulanacak kampanya ve promosyonlarda bu ilişkilerden yararlanmak, dolayısıyla satış gelirini arttırmaktır.

Apriori Algoritması'nın İşleyişi

Apriori Algoritması'nda öncelikle ürünlerin frekansları bulunarak Eşitlik (1) yardımıyla destek değerleri hesaplanmıştır. Örneğin motor yağı için destek değeri; motor yağının frekansının toplam müşteri sayısına bölünmesiyle bulunur ($17/40 = 0,425$). Daha sonra uzman görüşüne göre 0,1 olarak belirlenen minimum destek ölçütü ile her bir ürüne ilişkin hesaplanan destek değerleri karşılaştırılmıştır. Böylelikle destek değerleri minimum destek ölçütünden büyük veya bu değere eşit olan tekli ürünler belirlenmiştir. En fazla kaç ürünün birlikte ele alınabileceğini belirlemek amacıyla işlemlere devam edildiğinde bu sayının dört olduğu görülmüştür. Tablo 2'de dörtlü ürün gruplarından minimum destek değerini geçen ürünler yer almaktadır.

Tablo 2. Destek Sayıları Minimum Destek Sayısından Büyük Olan Dörtlü Ürünler

Ürünler	Frekans	Destek
motor yağı, fren balatası, hava filtresi, karter yağı boşaltma tapası	4	0,1
motor yağı, hava filtresi, karter yağı boşaltma tapası, yağ filtresi	6	0,15

Sonuç 1:

Elde edilen {motor yağı, fren balatası, hava filtresi, karter yağı boşaltma tapası} kümesini göz önüne alırsak örneğin; motor yağı, fren balatası, hava filtresi

→ karter yağı boşaltma tapası kuralı için güven ölçütü şu şekilde elde edilir:

güven(motor yağı, fren balatası, hava filtresi → karter yağı boşaltma tapası) =

$$\frac{\text{sayı}(\text{motor yağı, fren balatası, hava filtresi, karter yağı boşaltma tapası})}{\text{sayı}(\text{motor yağı, fren balatası, hava filtresi})} = \frac{4}{6} = \%66,67 \quad (3)$$

Birinci ürün grubu için tüm güven ölçütleri Tablo 3'teki gibidir.

Tablo 3. Sonuç 1 için Frekans ve Güven Düzeyi

Ürünler	Frekans	Güven (%)
motor yağı	17	23,5
fren balatası	11	36,4
hava filtresi	15	26,7
karter yağı boşaltma tapası	12	33,3
motor yağı, fren balatası	4	100,0
motor yağı, hava filtresi	14	28,6
motor yağı, karter yağı boşaltma tapası	12	33,3
fren balatası, hava filtresi	5	80,0
fren balatası, karter yağı boşaltma tapası	4	100,0
hava filtresi, karter yağı boşaltma tapası	11	36,4
motor yağı, fren balatası, hava filtresi	4	100,0
motor yağı, fren balatası, karter yağı boşaltma tapası	4	100,0
motor yağı, hava filtresi, karter yağı boşaltma tapası	11	36,4
fren balatası, hava filtresi, karter yağı boşaltma tapası	4	100,0

Tablodan, sadece motor yağı alanların motor yağının yanında fren balatası, hava filtresi, karter yağı boşaltma tapasını da alma olasılıkları %23,5 iken, fren balatası alanların aynı zamanda motor yağı, hava filtresi ve karter yağı boşaltma tapasını da alma olasılıkları %36,4, hava filtresi alanların aynı zamanda motor yağı, fren balatası ve karter yağı boşaltma tapasını alma olasılıkları %26,7; sadece karter yağı boşaltma tapasını alanların bu ürünün yanında motor yağı, fren balatası ve hava filtresini alma olasılıkları %33,3 çıkmaktadır. Motor yağı ve fren balatası alanlar %100 olasılıkla aynı zamanda hava filtresi ve karter yağı boşaltma tapası da almakta iken motor yağı ve hava filtresi alanların aynı zamanda fren balatası ve karter yağı boşaltma tapası da alma olasılıkları %28,6' dır. Motor yağı ve karter yağı boşaltma tapası alanların aynı zamanda hava filtresi ve fren balatası alma olasılıkları %33,3, fren balatası ve hava filtresi alanların motor yağı ve karter yağı boşaltma tapasını da beraberinde alma olasılıkları %80' dir. Fren balatası ve

karter yağı boşaltma tapası alanlar ise %100 olasılıkla motor yağı ve hava filtresi de almaktadırlar. Hava filtresi ve karter yağı boşaltma tapası alanlar %36,4 olasılıkla beraberinde motor yağı ve fren balatası da almaktayken, motor yağı, fren balatası ve hava filtresi alanların %100'ü karter yağı boşaltma tapası da almaktadırlar. Benzer şekilde motor yağı, fren balatası ve karter yağı boşaltma tapası alanlar da %100 olasılıkla hava filtresi almaktadırlar. Motor yağı, hava filtresi ve karter yağı boşaltma tapası alanlar %36,4 olasılıkla fren balatası almaktayken, fren balatası, hava filtresi ve karter yağı boşaltma tapası alanlar %100 olasılıkla motor yağı da almaktadırlar.

Sonuç 2:

Diğer ürün grubu olan {motor yağı, hava filtresi, karter yağı boşaltma tapası, yağ filtresi} kümesini göz önüne alırsak örneğin; motor yağı, hava filtresi, karter yağı boşaltma tapası → yağ filtresi kuralı için güven ölçütü şu şekilde elde edilir:

$$\text{güven}(\text{motor yağı, hava filtresi, karter yağı boşaltma tapası} \rightarrow \text{yağ filtresi}) = \frac{\text{sayı}(\text{motor yağı, hava filtresi, karter yağı boşaltma tapası, yağ filtresi})}{\text{sayı}(\text{motor yağı, hava filtresi, karter yağı boşaltma tapası})} = \frac{6}{11} = \%54,5 \quad (4)$$

İkinci ürün grubu için de tüm güven ölçütleri Tablo 4'teki gibi elde edilmiştir.

Tablo 4. Sonuç 2 İçin Frekans ve Güven Düzeyi

Ürünler	Frekans	Güven (%)
motor yağı	17	35,3
hava filtresi	15	40,0
karter yağı boşaltma tapası	12	50,0
yağ filtresi	9	66,7
motor yağı, hava filtresi	14	42,9
motor yağı, karter yağı boşaltma tapası	12	50,0
motor yağı, yağ filtresi	9	66,7
hava filtresi, karter yağı boşaltma tapası	11	54,5
hava filtresi, yağ filtresi	7	85,7
karter yağı boşaltma tapası, yağ filtresi	7	85,7
motor yağı, hava filtresi, karter yağı boşaltma tapası	11	54,5
motor yağı, hava filtresi, yağ filtresi	7	85,7
motor yağı, karter yağı boşaltma tapası, yağ filtresi	7	85,7
hava filtresi, karter yağı boşaltma tapası, yağ filtresi	6	100,0

Bu tablonun sonuçları ise şöyledir: Motor yağı satın alanların yanında hava filtresi, karter yağı boşaltma tapası ve yağ filtresi de alma olasılıkları %35,3, hava filtresi alanların yanında motor yağı, karter yağı boşaltma tapası ve yağ filtresini de alma olasılıkları %40, karter yağı boşaltma tapası alanların aynı zamanda motor yağı, hava filtresi ve yağ filtresini de alma olasılıkları %50, yağ filtresi alanların beraberinde motor yağı, hava filtresi, karter yağı boşaltma tapası da alma olasılıkları ise %66,7' dir. Motor yağı ve hava filtresi satın alanların aynı zamanda karter yağı boşaltma tapası ve yağ filtresini de alma olasılıkları %42,9, motor yağı ve karter yağı boşaltma tapası satın alanların aynı zamanda hava filtresi ve yağ filtresini de alma olasılıkları %50, motor yağı ve yağ filtresi satın alanların aynı zamanda hava filtresi ve karter yağı boşaltma tapası da alma olasılıkları %66,7' dir. Hava filtresi ve karter yağı boşaltma tapası alanların beraberinde motor yağı ve yağ filtresi alma olasılıkları %54,5 iken, hava filtresiyle yağ filtresi alanların motor yağı ve karter yağı boşaltma tapası da alma olasılıkları %85,7' dir. Karter yağı boşaltma tapası ile yağ filtresi satın

alanların beraberinde motor yağı ve hava filtresi de satın alma olasılıkları %85,7, motor yağı, hava filtresi ve karter yağı boşaltma tapası alanların yağ filtresini de beraberinde alma olasılıkları %54,5' tir. Motor yağı, hava filtresi ve yağ filtresi alanların karter yağı boşaltma tapasını da beraberinde alma olasılıkları %85,7, motor yağı, karter yağı boşaltma tapası, yağ filtresi alanların hava filtresini de alma olasılıkları %85,7, hava filtresi, karter yağı boşaltma tapası ve yağ filtresi alanların aynı zamanda motor yağını da beraberinde alma olasılıkları ise %100' dür.

FP-Growth Algoritması'nın İşleyişi

FP-Growth Algoritması'nın uygulaması için öncelikle çalışma kapsamında ele alınan 40 faturada yer alan ürünler için hesaplanan destek değerleri incelenerek 0,1'den büyük değere sahip olanlar belirlenir. Söz konusu ürünlerin frekans değerleri bulunur. Frekansı 1'den küçük olan ürünler FP-Growth Algoritması işlemlerine dahil edilmez. Tablo 5'te 40 fatura içinde destek değeri 0,1'den büyük olan nesnelere içinde frekans değeri 1'den küçük olanlar çıkarıldıktan sonra birlikte satın alınan nesnelere listesi yer almaktadır.

Tablo 5. FP-Growth Algoritması İçin Yaygın Nesne Kümeleri

motor yağı	karter yağı boşaltma tapası	fren balatası	hava filtresi	
fren balatası	şanzıman			
motor yağı	denge çubuk burçları	hava filtresi		
motor yağı	karter yağı boşaltma tapası	hava filtresi	yağ filtresi	
motor yağı	karter yağı boşaltma tapası	fren balatası	hava filtresi	denge çubuk burçları
motor yağı	karter yağı boşaltma tapası	fren balatası	hava filtresi	yağ filtresi
motor yağı	karter yağı boşaltma tapası	hava filtresi	yağ filtresi	
motor yağı	karter yağı boşaltma tapası	hava filtresi	yağ filtresi	
motor yağı	karter yağı boşaltma tapası	hava filtresi	yağ filtresi	
motor yağı	karter yağı boşaltma tapası	hava filtresi		
fren balatası	denge çubuk burçları	hava filtresi		
motor yağı	yağ filtresi	hava filtresi		
motor yağı	karter yağı boşaltma tapası	fren balatası	hava filtresi	
motor yağı	karter yağı boşaltma tapası	hava filtresi		
fren balatası	denge çubuk burçları			
motor yağı	yağ filtresi			
motor yağı	hava filtresi			
motor yağı	karter yağı boşaltma tapası	hava filtresi	yağ filtresi	

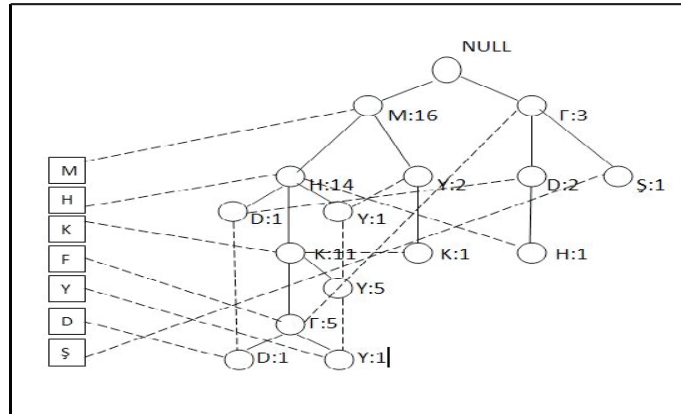
Ürünlerin frekans değerlerine göre büyükten küçüğe doğru sıralandığı “Başlık Tablosu” (Tablo 6) aşağıdaki gibidir.

Tablo 6. Başlık Tablosu

Ürünler	Frekansları
motor yağı	16
hava filtresi	15
karter yağ boşaltma tapası	12
yağ filtresi	9
fren balatası	7
denge çubuk burçları	4
şanzıman	1

Boş (null) kökten başlamak üzere “Başlık Tablosu”nda yer alan ürünlerin frekanslarına göre çizilen FP-Tree Şekil 1’deki gibidir. Şekilde M, motor yağını; H, hava filtresini; K, karter yağı boşaltma tapasını; F, fren balatasını; Y, yağ filtresini; D, denge çubuk burçlarını; Ş ise şanzımanı göstermektedir.

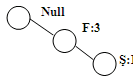
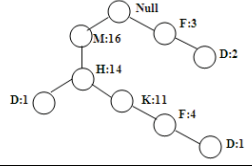
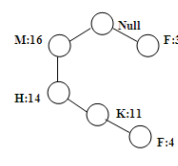
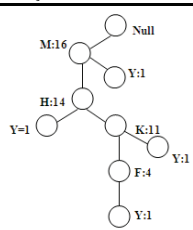
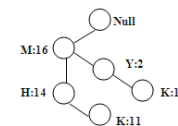
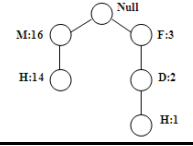
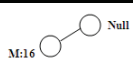
Birlikte sık alınan ürünleri bulmak için Şekil 1’de verilen FP-Tree üzerinden FP-Growth Algoritması işlemler yapılır. Algoritma en küçük frekanslı üründen başlar ki



Şekil 1. FP-Ağacı

bu da “Başlık Tablosu”ndaki en son ürün olan “Ş” dir. Daha sonra kökten ürüne kadar olan tüm dallar belirlenir ve her bir ürün için koşullu örüntüler ortaya

çıkartılır. “Başlık Tablosu”nda sıralanan her bir ürün için oluşturulan koşullu örüntüler Şekil 2’de yer almaktadır.

<i>S için:</i>	<i>D için:</i>
FŞ(1) 	MHKFD(1) MHD(1) FD(2) 
<i>F için:</i>	<i>Y için:</i>
MHKF(4) F(3) 	MHKFY(1) MHKY(5) MHY(1) MY(2) 
<i>K için:</i>	<i>H için:</i>
MHK(11) MYK(1) 	MH(14) FDH(1) 
<i>M için</i>	
M(16) 	

Şekil 2. Başlık Tablosu'nda Yer Alan Ürünler İçin Koşullu Örüntüler

Apriori Algoritması'nda olduğu gibi FP-Growth Algoritması sonucunda da oluşturulan koşullu örüntülere göre frekans değeri 1'den büyük olmak üzere birlikte alınan ürünlerin en fazla dörtlü gruplar oluşturduğu belirlenmiştir. Frekans değeri 1'den büyük olan dörtlü ürün gruplarından ise {motor yağı, hava filtresi, karter yağı boşaltma tapası, yağ filtresi}'nin en yüksek frekansa sahip ürün grubu olduğu belirlenmiştir.

Sonuç

Bu çalışmada Türkiye'de otomotiv sektöründe faaliyet gösteren bir yetkili servisin müşterilerine ait alış-veriş verileri, birliktelik kurallarından Apriori ve FP-Growth Algoritmaları kullanılarak analiz edilmiş ve müşterilerin hangi ürünleri birlikte satın aldıkları gözlemlenmiştir.

Buna göre Apriori Algoritması ile {motor yağı, fren balatası, hava filtresi, karter yağı boşaltma tapası} ve {motor yağı, hava filtresi, karter yağı boşaltma tapası,

yağ filtresi} olmak üzere dörtlü ürün grubundan oluşan iki sonuç kümesi elde edilmişken aynı veri kümesi için uygulanan FP-Growth Algoritması sonucunda ise sadece {motor yağı, hava filtresi, karter yağı boşaltma tapası, yağ filtresi} sonuç kümesi elde edilmiştir. Başka bir ifade ile ele alınan ürün gruplarının birliktelik kurallarını belirlemede Apriori Algoritması iki farklı çözüm verirken FP-Growth Algoritması ise tek bir çözüm vererek servis yetkililerinin ürün birlikteliklerini daha net bir şekilde değerlendirmeleri sağlanmıştır.

Çalışma sadece araştırmanın yapıldığı serviste yer alan ürünler ile gerçekleştirilmiştir. Elde edilen sonuçlar mevsim şartlarına göre otomobillerin ihtiyacı olan ürünlerin değişiklik gösterebilmesinden dolayı araştırmanın yapıldığı zaman aralığına göre farklılık gösterebilecektir. Bu durum servis yetkililerinin gerçekleştirecekleri kampanya ve promosyonlara da yansiyacaktır.

Kaynakça

- Agrawal, R., Imielinski, T. ve Swami, A. (1993).** Mining Association Rules Between Sets of Items in Large Databases. *In ACM SIGMOD Conf. Management of Data.*
- Agrawal, R. ve Srikant, R. (1994).** Fast Algorithms for Mining Association Rules. *20. VLDB Conference, Chile.* s.487-499
- Akpınar, H. (2000).** Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İ.Ü. İşletme Fakültesi Dergisi.* C:29, 1-22 ss.
- Alpaydın, E. (2000).** Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri. *Bilişim 2000 Veri Madenciliği Eğitim Semineri, İstanbul.*
- Berry, M.J.A ve Linoff, G.S. (2004).** Data Mining Techniques: For Marketing, Sales and Customer Relationship Management. *New York: John Wiley & Sons Inc.*
- Daniel H. (2011).** Performance Comparison of Apriori and FP-Growth Algorithms in Generating Association Rules. *Proceedings of the European Computing Conference, ISBN: 978-960-474-297-4.*
- Györödi, C., Györödi, R. ve Holban, S. (2004).** A Comparative Study of Association Rules Mining Algorithms. *SACI 2004, 1st Romanian-Hungarian Joint Symposium on Applied Computational Intelligence, Timisoara, Romania, 213-222.*
- Han, J. ve Kamber, M. (2000).** *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, Burnaby.
- Han, J., Pei, H. ve Yin, Y. (2000).** *Mining Frequent Patterns without Candidate Generation*, SIGMOD'00, ACM Press, New York, NY, USA.
- Özkan, Y. (2008).** *Veri Madenciliği Yöntemleri*, 1. basım, Papatya Yayınları.
- Pandey, A., ve Pardasani, K.R., (2009).** Rough Set Model for Discovering Multidimensional Association Rules. *International Journal of Computer Science and Network Security (IJCSNS)*, 9(6): 59-164.
- Silahtaroglu, G. (2008).** *Kavram ve Algoritmalarıyla Veri Madenciliği*, 1. Basım, Papatya Yayınları.
- Surendiran R., Rajan.K.P. ve Sathish K.M. (2010).** Study on the Customer Targeting Using Association Rule Mining. *Surendiran et. al./ (IJCSSE) International Journal on Computer Science and Engineering*, Vol. 02, No. 07, 2010, 2483-2484, ISSN : 0975-3397.
- Wei Z., Hongzhi L. ve Na Z. (2008).** Research on the FP Growth Algorithm about Association Rule Mining. *Business and Information Management, ISBIM '08. International Seminar on*, pp. 315-318, ISBN: 978-0-7695-3560-9.
- Vishal S., Nikita J. ve Sharad V., (2010).** Efficient Use of Apriori Algorithm for Accessing Supermarket Database. *IT and Business Intelligence. Proceedings of 2nd international Conference on IT & Business Intelligence (ITBI-10)*, Organized by Institute of Management Technology, Nagpur, India and Technically Sponsored by IEEE CIS ISBN No: 978-81-7446-900-7.