



## Veri Madenciliğinde Birliktelik Kuralları ve İkinci El Otomobil Piyasası Üzerine Bir Uygulama

*Association Rules in Data Mining and an Application in Second Hand Car Market*

**Mehmet ÖZÇALICI<sup>1</sup>**

**Geliş Tarihi:** 19.12.2016 / **Düzenleme Tarihi:** 26.03.2017 / **Kabul Tarihi:** 27.03.2017

### Özet

İkinci el piyasada devredilen otomobil sayısı her geçen yıl artmaktadır. Bu durum ikinci el araç piyasasının detaylı bir şekilde incelenmesini zorunlu kılmaktadır. Yeni geliştirilen analizler ve yığınla toplanabilen veri setleri yardımıyla ikinci el araç piyasasının işleyiş mekanizmasına ilişkin faydalı bilgiler ortaya çıkabilmektedir. Çalışmanın amacı, ikinci el araçlara ilişkin verilerin bulunduğu bir veri tabanını birliktelik kuralları ile incelemektir. Birliktelik kuralları bir arada gözlenen özelliklerin belirlenmesinde başarıyla kullanılmaktadır. Bu çalışma için, 2016 yılının Temmuz ayında ve Ağustos ayının ilk üç haftasında ikinci el otomobil ilanlarının yer aldığı bir siteden 211109 adet otomobile ait 73 adet değişken web kazıma tekniği ile bir araya getirilmiştir. Birliktelik kurallarının oluşturulabilmesi için veri setinin mantıksal (boolean) değişkenlerden oluşması gerekmektedir. Sayısal değişkenler, dağılımları göz önünde bulundurulmak üzere mantıksal yapıya çevrilmiştir. Çalışmada apriori algoritması ile oluşturulan birliktelik kuralları ağ grafiği yardımı ile görselleştirilmiştir. Çalışma sonunda “dizel araçlar az yakar, fazla hız yapamazlar, torkları yüksektir; yeni ve pahalı araçların vergisi yüksektir” gibi kurallar oluşturulmuştur. Çalışmanın sonuçları ikinci el araç piyasasında işlem yapan taraflar için yol gösterici nitelikte bilgiler içermektedir.

**Anahtar Kelimeler:** Veri madenciliği, apriori algoritması, ikinci el oto, birliktelik kuralları, ağ grafiği, web kazıma.

### Abstract

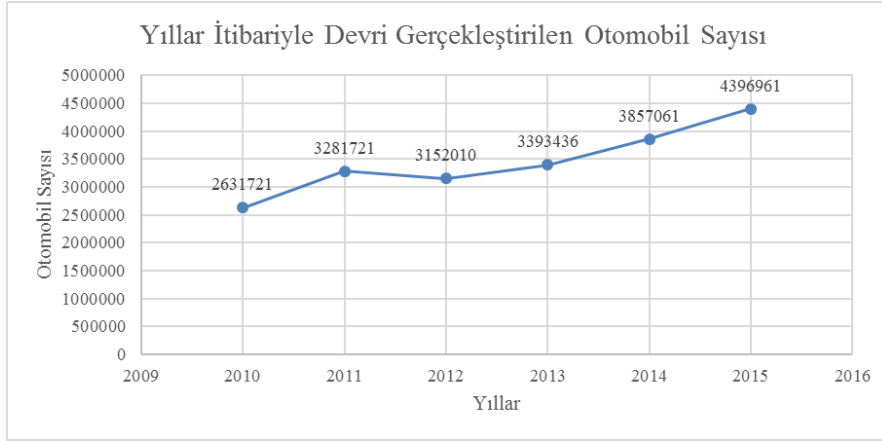
Number of road motor vehicles that are handed over are increasing over the years. That makes the detailed examination of second hand car market is a necessity. It is possible to extract useful information about the operating mechanism of second hand car market using new analysis techniques and massive datasets. The purpose of this study is to examine the database which includes data belong to second hand cars by association rules. Association rules are successful at determining which two objects are observed simultaneously in a dataset. For this study, a dataset includes to 21109 second hand cars with 73 variables is scraped from a website in July 2016 and August 2016. In order to apply the apriori algorithm, the dataset must be consisted from Boolean variables. Numerical variables are converted to Boolean type by using their distribution. The association rules that are created with apriori algorithm is visualized with a network graph. At the end of the study rules such as “diesel cars have less fuel consumption, cannot go much faster, have higher torque values; new and expensive cars have higher taxes” are created. The overall results have useful information for those who operate in second hand car market.

**Keywords:** Data mining, apriori algorithm, second-hand car, association rules, network graphs, web scraping.

## 1. Giriş

Otomobiller, hanehalkı için gayri menkullerden sonra ikinci en çok satın alınan (veya alınmak istenen) bir varlıktır. Yapılan araştırmalara göre bazı ülkelerde ikinci el otomobil satış hacimlerinin birinci el satış hacminden daha yüksek düzeye çıktığı belirlenmiştir (Asılkan ve İrmak, 2009). Türkiye İstatistik Kurumu'ndan edinilen verilerle oluşturulan Şekil 1'de yıllar itibarıyla devir gerçekleştirilen otomobil sayısının artmakta olduğu görülmektedir. Bu durum ikinci el araç piyasasının ülke ekonomisi içindeki yerinin detaylı bir şekilde incelenmesi gereken bir boyutta olduğunu ortaya çıkarmaktadır.

<sup>1</sup>Yrd. Doç. Dr., Kilis 7 Aralık Üniversitesi İİBF, Uluslararası Ticaret ve Lojistik Bölümü. Kilis, Türkiye.  
E-Posta: mozclici@gmail.com



**Şekil 1:** Yıllar itibariyle devri gerçekleştirilen otomobil sayısı (Kaynak: [http://www.tuik.gov.tr/PreIstatistikTablo.do?istab\\_id=2248](http://www.tuik.gov.tr/PreIstatistikTablo.do?istab_id=2248) Erişim tarihi : 30.11.2016)

Çok fazla alıcı ve satıcının işlem gördüğü ikinci el otomobil piyasasında satışa sunulan araçların genel özelliklerinin bilinmesi alım ve satım işlerinin daha rasyonel bir şekilde gerçekleştirilmesine neden olacak ve bu durum da ulusal kaynakların etkin bir şekilde kullanılmasını sağlayacaktır. Ayrıca araçlar hakkında uzman bilgisine sahip olmayan alıcı ve satıcılar da piyasadaki araçların genel durumunu göz önünde bulundurmak suretiyle daha sağlıklı kararlar alabilirler.

Türkiye’de ikinci el araçların satış fiyatını tahmin etmeye yönelik çalışmalar gerçekleştirilmiştir (Daştan, 2016; Asilkan ve İrmak, 2009; Ecer, 2003). Bu çalışmalarda ikinci el otomobil fiyatının belirleyicilerine ilişkin faydalı bilgiler ortaya çıkmaktadır. Bununla birlikte, piyasada satışa sunulan araçların sahip olduğu özellikler veri madenciliği teknikleri ile henüz incelenmemiştir. Farklı markaların bir arada olduğu ikinci el piyasada yer alan araçların özelliklerinin birliktelik kuralları ile incelenmesi mümkündür.

Bu çalışmanın amacı Türkiye ikinci el araç piyasasında satışa sunulan araçlara ait genel bir durum değerlendirmesi yapmaktır. Veri madenciliği tekniklerinden olan apriori algoritması piyasada satışa sunulan araçlara ilişkin kuralların oluşturulmasında kullanılmıştır ayrıca kurallar ağ grafiği ile görselleştirilmiştir. Apriori algoritmasının ikinci el otomobil piyasasından elde edilen veri setine uygulanması neticesinde, hangi özelliklerin piyasadaki araçların çoğunda birlikte yer aldığı bilgisini açığa çıkarmaktadır.

Çalışma beş bölümden oluşmaktadır. Bu giriş bölümünden sonra, ikinci bölümde literatür araştırması yer almaktadır. Üçüncü bölümde veri madenciliği ve apriori algoritması açıklanacaktır. Dördüncü bölümde çalışmada kullanılan veri seti tanıtılacak ve analiz sonuçlarına yer verilecektir. Son bölümde ise sonuç tartışma ve gelecek çalışmalar için öneriler yer alacaktır.

## 2. Literatür Taraması

Apriori algoritması ile kural oluşturma, temel veri madenciliği görevlerinden biridir. Apriori algoritmasının literatürde çok farklı amaçlar için kullanıldığı belirlenmiştir. Bu çalışmaları yayınladıkları yılları göz önünde bulundurmak suretiyle şu şekilde özetlemek mümkündür.

Kaur ve Kang (2016) apriori algoritmasının, zamanın dikkate alınmadığı statik veri setleri üzerinde çalıştığına dikkate çekmekte ve yeni bir algoritma önermektedirler. Önerdikleri algoritmada sadece statik özellikler değil, zamanla değişen özellikler de dikkate alınmaktadır.

Doğrul vd (2015) yaptıkları çalışmada trafik kazası verilerini analiz etmek için birliktelik kurallarından faydalanmışlardır. Çalışmada kazaların sıklıkla yaşandığı yerler ve zamanlar belirlenmiştir ve bu gibi durumlarda önlemlerin artırılması ile birlikte kazaların önlenebileceği ifade edilmektedir. Kurt Pehlivanoğlu ve Duru (2015) çalışmalarında bir ortaokulda okuyan öğrencilerin sosyal medya kullanma alışkanlıkları üzerine veri toplamış ve sonuçları birliktelik kuralları yardımıyla analiz etmişlerdir. Çalışmanın sonucunda ortaya çıkan kurallar raporlanmaktadır.

Doğan vd (2014) yaptıkları çalışmada bir sigorta şirketinin müşterilerine ilişkin toplanan bir veri seti üzerinde birliktelik kurallarını uygulamışlardır. Çalışma sonunda birlikte satın alınan sigorta çeşitleri belirlenmiş ve pazarlama faaliyetlerine yön gösterebilecek bilgilerin ortaya çıktığı raporlanmaktadır. Keleş ve Kaya (2014) ise yaptıkları çalışmada duvar işçiliğinde verimliliği etkileyen faktörleri apriori algoritması ile incelemişlerdir.

Aytaç ve Bilge (2013) yaptıkları çalışmada tele pazarlama verilerini kullanmış ve birliktelik kuralları oluşturmuşlardır. Çalışmalarında sayısal değişkenler, apriori algoritmasının işleyebilmesi için kategorik hale getirilmiştir. Çalışma sonunda kurallar oluşturulmuş ve pazarlama stratejileri için değerli olan bazı bilgilerin ortaya çıktığı ifade edilmektedir. Erpolat (2012) yaptığı çalışmada otomobil yetkili servislerinde apriori ve fast grow algoritmasını kullanmak suretiyle birlikte satın alınan ürünleri belirlemiştir.

Ay ve Çil (2008) yaptıkları çalışmada bir perakende zincirinde birliktelik kuralları oluşturmuşlardır. Çalışmada oluşturulan birliktelik kurallarından yola çıkmak suretiyle aynı zamanda yerleşke düzeni de önerilmektedir. Timor ve Şimşek (2008) yaptıkları çalışmada tüketici davranışlarını modellemek için sepet analizi / birliktelik kurallarını kullanmışlardır. Büyük bir perakende zincirine ait veriler kullanılmıştır. Çalışma sonunda elde edilen birliktelik kurallarının pazarlama ve raf diziminde faydalı olabileceği raporlanmaktadır. Erdem ve Özdağoğlu (2008) yaptıkları çalışmada acil hasta verilerine apriori algoritmasını uygulamışlardır. Hastalara ilişkin anlamlı kurallar ortaya çıkmıştır.

Cavique (2007) ise çalışmasında birliktelik kuralı bulma algoritmalarının bazılarının büyük veri setlerinde çok uzun zamanda sonuçlanacağını vurgulamaktadır. Çalışmalarında büyük veri setlerinde hızlı bir şekilde birliktelik kurallarını bulacak bir algoritma tanıtılmaktadır. Çalışmalarında önerdikleri yöntemin başarılı olduğu vurgulanmaktadır. Özçakır ve Çamurcu (2007) yaptıkları çalışmada apriori algoritmasını kullanan bir yazılım tasarımı gerçekleştirmişlerdir. Bir firmanın pastane satış verileri üzerinde geliştirdikleri yazılımı çalıştırmışlardır. Genelde aynı nitelikteki ürünlerin birlikte satıldığı sonucuna varmışlardır.

Karabatak ve İnce (2004) öğrencilerin başarılarını analiz etmek için apriori algoritmasından yararlanmıştır ve bazı kurallar oluşturmuşlardır. Çalışma sonunda üniversiteye yerleşme puanları ve ilk dönem ders notları ile sonraki dönem ders notları arasında dikkate alınması gereken ilişkiler bulunmuştur.

Türkiye'deki ikinci el otomobil piyasasında ise araçların satış fiyatlarını tahmin etmeye yönelik çalışmalar da mevcuttur. Örneğin, Daştan (2016) çalışmasında araçlara ilişkin marka, model, yaş, yakıt türü, vites, kilometre ve benzeri özelliklerin aracın fiyatına olası etkilerini araştırmıştır. Yazar çalışmasında çeşitli internet sitelerinde Ekim 2015 tarihinde yer alan ilanlardan 1000 tanesini kullanmıştır. Hedonik modeller kurulmuş ve en başarılı modelin yarı logaritmik model olduğu sonucunu raporlanmaktadır. Asilkan ve Irmak (2009) çalışmasında Avrupa'da 40 adet web sitesinden 3 yıl boyunca sistematik olarak topladığı verileri kullanmıştır. Yazarlar çalışmasında en popüler 3 adet otomobil modelinin 1 yaşındaki ve 20 000 km'deki ilanlarının, Ocak 2005 – Aralık 2007 tarihleri arasındaki 3 yıllık günlük fiyat verilerinin aylık ortalamalarından oluşan zaman serilerini kullanmıştır. Yapay sinir ağları ve zaman serileri analizi yöntemlerini kullanmak suretiyle fiyatı tahmin etmiştir. Sonuçta yapay sinir ağlarının daha başarılı bir şekilde fiyat tahmini gerçekleştirebildiği raporlanmaktadır. Ecer (2003)'ün çalışmasında 2012 yılı içerisinde Türkiye'de faaliyet gösteren iki adet internet sitesinde yer alan ilanlardan rasal olarak seçilen 640 adet otomobile ilişkin veriler kullanılmıştır. Araçların yayınlanan özellikleri tahmin modellerinde girdi olarak kullanılmıştır. Yapay sinir ağlarının hedonik modele göre daha yüksek performans sergilediği raporlanmaktadır. Bununla birlikte Akçi (2016) çalışmasında tüketicilerin ikinci el otomobillere bakış açısını incelemiştir. 271 adet müşteri ile anket çalışması gerçekleştirilmiş ve müşterilerin en çok arkadaşlarına ve tanıdıklarına güvendikleri sonucunu raporlamıştır.

Türkiye'de yer alan çalışmalar incelendiğinde ikinci el otomobillerin özelliklerinin birliktelik kuralları yardımıyla incelendiği bir çalışmaya rastlanmamıştır. Bu çalışma aynı zamanda literatürdeki bu açığı kapatmayı da amaçlamaktadır.

### 3.Yöntem

#### 3.1. Veri Madenciliği

Veri madenciliği veri setinde var olan anlamlı örüntülerin (patterns) keşfedilmesi sürecidir (Witten vd, 2011:5). Veri madenciliği bu örüntüleri çeşitli algoritmalar yardımıyla keşfetmektedir. Bir veri madenciliği algoritması ise, veriyi girdi olarak kabul eden ve örüntü şeklinde çıktı üreten iyi tanımlanmış prosedür olarak tanımlanabilir (Hand vd 2001:141). Veri madenciliği, bankacılık, telekomünikasyon ve sigorta endüstrisinde hileli işlemlerin tespitinde, telefon ve kredi kartı işletmelerinde kaybedilme olasılığı olan müşterilerin belirlenmesinde (churn management), üretim işletmelerinde kalite kontrolü için ve benzeri birçok alanda uygulanmaktadır (Olson ve Delen, 2008:4).

Veri madenciliğinin en popüler üç adet algoritması Sınıflandırma ve Regresyon Ağaçları (CART), geri yayılım (Backpropagation) ve apriori algoritmasıdır ve bu algoritmalar sırasıyla sınıflandırma, regresyon ve kural keşfi için kullanılmaktadır (Hand vd 2001:143). Apriori algoritması ile mantıksal (boolean) birliktelik kurallarını keşfetmek mümkündür (Han ve Kamber, 2006:235). Apriori algoritmasının işletmecilik alanındaki en yaygın uygulama alanı pazar sepeti analizi olarak bilinmektedir. Pazar sepeti analizi ile birlikte, müşterilerin satın alma alışkanlıkları incelenmektedir (Han ve Kamber, 2006:272). Pazar sepeti analizi sonucunda sıklıkla birlikte satılan ürünler belirlenmektedir ve buna bağlı olarak pazarlama stratejileri şekillenmektedir. Özellikle e-ticaretin gelişmesi ile birlikte internet sitelerinde alışveriş yaparken sıklıkla rastlanabilen, "bunu alan müşteriler bunu da tercih etti" veya "ilgili ürünler" gibi başlıklar altında benzer ürünlerin listelenebilmesi apriori algoritması yardımıyla mümkün hale gelmektedir. Bunun haricinde apriori algoritması öğrenci başarı analizi (Karabatak ve İnce, 2004), sigortacılık sektöründe müşteri ilişkileri yönetimi (Doğan vd, 2014), acil hasta verilerinin analizi (Erdem ve Özdağoğlu, 2008), trafik kazalarının analizi (Doğrul vd. 2015; Pande ve Abdel-Aty, 2009) gibi farklı alanlarda da başarı ile uygulanmaktadır.

#### 3.2. Birliktelik Algoritmaları

İlişki kuralları bilgisayar bilimleri alanında geliştirilmiş bir analiz türüdür (Giudici ve Figini, 2009:90). İlişki kural madenciliğinde (Association Rule Mining) sıklıkla birlikte bir arada ortaya çıkan ilişkilerin  $X \rightarrow Y$  formatında açığa çıkarılması sağlanmaktadır. Bu ifade şu anlama gelmektedir:  $X$  'in ortaya çıktığı bir durumda büyük olasılıkla  $Y$  de ortaya çıkacaktır. İfadenin sol tarafındakilere (birden fazla olabilir) öncül(antecedent) denilmektedir ve sağ taraftakilere sonuç (ardıl veya descendent) denilmektedir. Eğer sol taraftaki ifade doğrudur, sağ taraftaki ifade de doğrudur (Hand vd 2001:428). Bu kuralın önem derecesi iki farklı ölçü ile ölçülebilir (Giudici ve Figini, 2009:92). Bunlar destek (support) ve güven (confidence) dir.

Destek değeri ile bütün veri setinin yüzde kaçının kuralı kapsadığı ölçülmektedir. Formül 1 yardımıyla hesaplanmaktadır:

$$destek = \frac{n(X \cup Y)}{N} \quad (1)$$

Formülde  $n(X \cup Y)$  ifadesi X ve Y nin birlikte yer aldığı işlem sayısını, N ise toplam işlem sayısını temsil etmektedir. Bu değer bir olması, incelenen veri setinde her işlemde X ve Y nin birlikte gerçekleştiğini, sıfır olması ise veri setinde hiçbir işlemde X ve Y nin birlikte gerçekleşmediğini ifade etmektedir.

Güven değerinde ise X i içeren işlemlerin yüzde kaçının Y yi de içerdiğini ifade etmektedir ve Formül 2 yardımıyla hesaplanmaktadır:

$$güven = \frac{n(X \cup Y)}{n(X)} \quad (2)$$

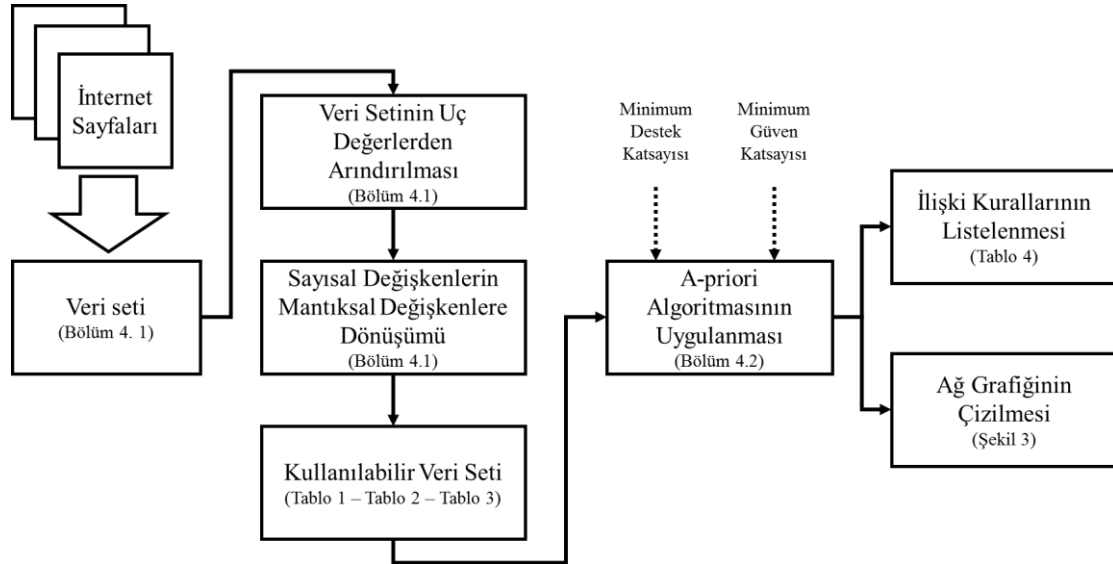
formülde destek formülünden farklı olarak paydada toplam X i içeren gözlem sayısı yer almaktadır. Bu değer bir çıkması X i içeren her işlemin Y yi de içerdiğini, sıfır çıkması ise X i içeren işlemlerin hiçbirinin Y yi içermediğini ifade etmektedir.

Apriori algoritmasının uygulanması için kullanıcı tarafından belirlenen minimum güven ve minimum destek değerlerinin belirlenmesi gerekmektedir. Bu minimum şartları sağlayan kurallar güçlü kurallar olarak adlandırılmaktadır (Han ve Kamber, 2006:230). Bazı kaynaklarda bu kurallar ilginç kurallar olarak da adlandırılmaktadır. Orta büyüklükteki bir veri setinde çok fazla sayıda kuralın üretilmesi olanaklıdır. Ancak veri setinin genelini kapsamayan kuralların incelemeye değer olmadığı da bir gerçektir. Bu nedenle, her kuralın değil de sadece belirli şartları sağlayan kuralların listelenmesi gerekmektedir. Minimum güven ve minimum destek değerlerinin kural sayısını kısıtladığını söylemek mümkündür. Destek ve güven değerlerinin en yükseği %100 (ya da 1) değerini almaktadır ve ilgili kuralın veri setinin tamamında yer aldığını ifade etmektedir.

Apriori algoritması pazarlama için faydalı bilgiler sunabilmektedir. Aynı zamanda çıktıları okumak kolaydır. Bu üstünlüklerine rağmen, fazla sayıda kural çıkarıldığında, kuralların anlaşılması zorlaşabilmektedir. Algoritmanın ikinci dezavantajı ise çok büyük veri setlerinde aşırı uzun hesaplama zamanına ihtiyaç duymasındır (Cavique, 2007).

#### 4. Analiz Ve Sonuçlar

Çalışmanın modeli Şekil 2'de yer almaktadır. Çalışma üç aşamadan oluşmaktadır. İlk aşamada veri seti toplanmıştır ve analize hazır hale getirilmiştir. İkinci aşamada analiz gerçekleştirilmiştir ve son olarak da analiz sonuçları kurallar ve ağ grafiği ile görselleştirilmiştir.



Şekil 2: Çalışmanın şematik çizimi

#### 4.1. Veri Seti

Çalışmada, ikinci el araç piyasasında satışa sunulan araçlara ilişkin bilgileri içeren bir veri setine ihtiyaç duyulmaktadır. İhtiyaç duyulan bu veri seti piyasada hazır olarak bulunmamaktadır. Bu nedenle yeni bir veri setinin bir araya getirilmesi gerekmektedir. Analizde kullanılan veri seti sahibinden.com adresinden web kazıma (web hasat veya web veri çekimi) yöntemi ile temin edilmiştir. Web kazıma tekniği, internet sitelerinden bilgisayar programı yardımıyla veri toplama işlemi olarak tanımlanabilir. Çalışmada sadece bir tek siteden faydalanılmasının nedeni, farklı sitelerde farklı nitelikteki bilgilerin yayınlanıyor olmasıdır. Başka bir ifade ile bir sitede yayınlanan araçlara ilişkin bir özellik bir diğer sitede yer almayabilmektedir. Bütünlük sağlamak adına sadece bir adet sitede yer alan ilanlar kullanılmıştır. Çalışmada kullanılan veri seti, ilgili sitede 2016 yılının Temmuz ayında ve Ağustos ayının ilk üç haftasında yayınlanan ilanlardan derlenmiştir. Toplamda 335287 adet ilan toplanmıştır. Ancak bir araya getirilen veri setindeki her ilan kullanılabilir nitelikte değildir. Bu nedenle veri setine temizleme işleminin (ön-ışleme) uygulanması gerekmektedir.

İlk olarak, aynı ilanların veri setinde yer alması ihtimaline karşı tarama yapılmış ve birbirinin aynı olan ilanlardan sadece birinin analizde yer alması sağlanmıştır. Bu işlemin sonucunda veri setinde 330860 adet otomobil ilanına ilişkin bilgiler kalmıştır. Bazı ilanlar uç değer niteliği taşıyabilir. Ayrıca bazı değişkenler ilan sahibi tarafından hatalı girilebilir. Örneğin veri setinde bir aracın modeli 2111 olarak girilmiştir. Bu gibi hatalı verilerin belirlenmesi ve veri setinden çıkarılması gerekmektedir. Bu amaçla sayısal değişkenler (fiyat, model yılı, kilometresi, motor hacmi, motor gücü, tork, şehir içi yakıt tüketimi, şehir dışı yakıt tüketimi, ortalama yakıt tüketimi ve taşıma kapasitesi) bazında veri seti küçükten büyüğe doğru sıralanmış ve en yüksek ve en düşük yüzde beşlik dilimde yer alan araçlar uç değer olarak nitelendirilmiştir. Yüzdeler dilimleri düşük belirlemek (%1 gibi) veri setinde uç değer olarak nitelendirilecek eleman sayısının düşük çıkmasına neden olacakken, bu değer yüksek belirlenmesi (%10 gibi) fazla sayıda elemanın uç değer olarak nitelendirilmesine neden olacaktır. Bu nedenle çalışma için %5 gibi orta düzeyde bir eşik değeri uygun kabul edilmiştir. Bu işlemin sonunda veri setinde analize uygun 211109 adet ilan kalmıştır. Çalışmada yer alan markalara ilişkin frekans dağılımı Tablo 1'deki gibidir. Tablo 1'e göre en çok yer alan ilanlar sırasıyla Renault (%16), Opel (%13), Volkswagen (%12) ve Ford (%10) marka arabalara aittir.

**Tablo 1: Çalışmada Kullanılan Araçların Markalara Göre Frekans Dağılımı**

Marka	Frekans	Yüzde	Marka	Frekans	Yüzde
Alfa-Romeo	943	0,4467	Mini	726	0,3439
Audi	3183	1,5078	Mitsubishi	599	0,2837
BMW	2876	1,3623	Nissan	1663	0,7877
Chevrolet	3832	1,8152	Opel	27536	13,0435
Chrysler	22	0,0104	Peugeot	11238	5,3233
Citroen	7416	3,5129	Renault	34117	16,1608
Dacia	1976	0,936	Seat	4200	1,9895
Daewoo	26	0,0123	Skoda	2860	1,3548
Daihatsu	74	0,0351	Smart	51	0,0242
Fiat	16720	7,9201	Subaru	121	0,0573
Ford	21404	10,1388	Suzuki	265	0,1255
Honda	8051	3,8137	Tata	175	0,0829
Hyundai	16258	7,7012	Toyota	11425	5,4119
Kia	2302	1,0904	Volkswagen	27209	12,8886
Mazda	823	0,3898	Volvo	1420	0,6726
Mercedes-Benz	1598	0,757	Toplam	211109	100

Tablo 2'de veri setinden bir kesit yer almaktadır. Bütün veri setini burada listelemek mümkün değildir. Okuyucunun veri setini anlayabilmesi açısından bir kesit sunulmuştur.

**Tablo 2: Veri Setinden Bir Kesit**

Sıra	ABS	ASR	ESP	EBD	...	YKD	Isofix	Alarm	M.Kilit
1	0	0	0	0	...	0	0	1	1
2	0	0	0	0	...	0	0	0	0
3	0	0	0	0	...	0	0	0	1
4	1	1	0	0	...	0	0	1	1
5	1	0	0	1	...	0	0	0	1
...	...	...	...	...	...	...	...	...	...
211105	1	0	0	0	...	0	1	0	0
211106	1	0	1	1	...	1	1	0	1
211107	1	0	1	1	...	0	1	1	1
211108	1	0	0	0	...	1	1	1	1
211109	0	0	0	0	...	0	0	0	0

Çalışmada sitede yer alan ve kullanıcıların belirleyebildiği bütün değişkenler kullanılmıştır. Kullanılan değişkenler araçların güvenlik, iç donanım, dış donanım, multimedya ve teknik özelliklerinden oluşmaktadır. Çalışmada kullanılan veri setinin tanımlayıcı istatistikleri Tablo 3'deki gibidir. Tablodaki değerler incelenirken, bazı bilgilerin kullanıcılar tarafından eksik girildiği (missing values) göz önünde bulundurulmalıdır.

Veri setinde bazı değişkenler ikili (ABS sistemi var veya yok) değişkenlerden oluşmaktadır ve bazı değişkenler de sayısal niteliklidir (aracın fiyatı, kilometresi gibi). Ancak apriori algoritmasının işleyebilmesi için veri setinin mantıksal değerlerden oluşması gerekmektedir. Bu nedenle sayısal olan değişkenler Tablo 3'deki gibi belirli sınıf aralıklarında mantıksal hale getirilmiştir. Sınıf aralıklarının belirlenmesinde her bir değişkenin kendi dağılımı göz önünde bulundurulmuştur. Her bir sayısal değişkenin en küçük, en yüksek, ortalama ve medyan değerleri hesaplanmış ve her gruba yaklaşık olarak eşit sayıda gözlem düşecek şekilde sınıf aralıkları belirlenmiştir.

**Tablo 3: Çalışmada Kullanılan Değişkenlere İlişkin Tanımlayıcı Bilgiler**

Sıra No	Değişken Adı	Doğru mu?	Sıra No	Değişken Adı	Doğru mu?
1	2000 < MY ≤ 2005	33642	38	Yıllık MTV > 200 TL	118556
2	2005 < MY < 2010	53691	39	ABS	183484
3	2010 < MY ≤ 2014	94744	40	ASR	59957
4	20000 < F ≤ 30000	46496	41	ESP	87775
5	30000 < F ≤ 50000	100768	42	EBD	93692
6	50000 < F ≤ 70000	40631	43	Yokuş Kalkış Dest.	56699
7	Benzinli	52801	44	ISOFIX	132148
8	Dizel Araç	100242	45	Alarm	62374
9	Benzin ve LPG	58042	46	Merkezi Kilit	182380
10	Manuel Vites	151195	47	Immobilizer	168779
11	Yarı Otomatik Vites	35929	48	Kumaş Koltuk	60946
12	20000 km den fazla	197254	49	Elkt Ön Cam	182804
13	1.4 Motor	79974	50	Elkt Arka Cam	130775
14	1.6 Motor	99836	51	Ön Kol Dayama	98568
15	M≤100 beygir	112723	52	Arka Kol Dayama	54168
16	M 100-150 beygir	94958	53	Hidrolik Dir.	119875
17	maxtorka≤= 200Nm	102964	54	Fonks. Dir.	76087
18	maxtorka> 200Nm	49370	55	Ayarlanabilir Dir	140432
19	maxtorkb≤=2500rpm	74903	56	Deri Direksiyon	69041
20	maxtorkb>2500rpm	54624	57	Soğutmalı Torpido	53324
21	0-100 10 s den fazla	140488	58	Sis Farı	149846
22	Azamisurat≤=200km/sa	139741	59	Elkt Ayna	165509
23	Ş. içi 100 km 5-10 litre	127423	60	Isıtmalı Ayna	113436
24	Ş. dışı 100 km 5 lt yakıt	94252	61	Arka Park Sensoru	60952
25	Ş. dışı ≥100 km 5 lt	57899	62	Alaşımlı Jant	131899
26	Ort. 100 km 5 lt 'den az	73817	63	Yağmur Sensoru	63864
27	Ort. 100 km 5 lt 'den fazla	78334	64	A. Cam Buz Önleyici	62676
28	Ykt Depo h≤=60 lt	137003	65	Kaset Çalar	151807
29	Uzunluk ≤= 4300 mm	68115	66	CD Çalar	167360
30	Uzunluk > 4300 mm	84036	67	MP3 Çalar	143106
31	Genişlik >1700 mm	119586	68	Navigasyon	74686
32	Yükseklik > 1400	148251	69	Telefon	85293
33	Net Ağırlık ≤= 1200	63187	70	USB bağlantısı	63805
34	Net Ağırlık > 1200	88583	71	AUX	70632
35	Taşıma kap. ≤= 500 kg	78682	72	IPOD bağlantısı	74243
36	Taşıma kap. > 500 kg	68255	73	Hoparlor	74570
37	Bagaj Kap. 200-500 lt	118245			

#### 4.2. Apriori Algoritmasının Uygulanması

Uygulama MATLAB platformunda gerçekleştirilmiştir. Algoritmanın çalışması için kullanıcının minimum destek ve minimum güven değerlerini belirlemesi gerekmektedir. Bu değerlerin yüksek tutulması düşük sayıda kuralın ortaya çıkmasına neden olmaktadır. Aynı şekilde bu değerlerin çok düşük tutulması da çok fazla sayıda kuralın ortaya çıkmasına neden olacaktır. Başka bir ifade ile minimum destek ve minimum güven değerlerinin kural sayısını kontrol ettiğini söylemek mümkündür. Destek ve güven değerleri bu çalışmada deneme yanılma yoluyla belirlenmiştir. Örneğin, minimum destek değeri %80 olarak belirlendiğinde sadece 12 adet kuralın oluşturulduğu görülmüştür. Yapılan denemeler sonucunda, minimum destek değeri %10 ve minimum güven değeri %70 olarak belirlendiğinde elde edilen kural sayısı ideal olarak değerlendirilmiştir.

Minimum destek ve güven seviyesinde hesaplanan her bir kural Tablo 4'de listelenmiştir. Tablo 4'de yer alan 1 numaralı kurala ilişkin destek değeri şu şekilde açıklanabilir. Bütün veri setinin %45.20 'sinde fiyatı 30 000 TL ile 50 000 TL arasında (30000 < F ≤ 50000) değişen araçlar 20 000 km'den fazla kullanılmıştır. Güven değeri ise şu şekilde açıklanabilir. Fiyatı 30 000 ile 50 000 TL arasında olan araçların %94.70 'i 20 000 km den fazla kullanılmıştır.

**Tablo 4. Çalışma Sonunda Elde Edilen Kurallar**

Kural No.	X	Y	Destek (%)	Güven (%)
1	F 30000 ile 50000	20000 km den fazla	45.2018	94.6977
2	Dizel Araç	20000 km den fazla	44.99	94.7487
3	F 30000 ile 50000	ABS	44.5291	93.2885
4	F 30000 ile 50000	Elektrikli Ön Cam	42.8248	89.718
5	MY 2010 ile 2014	20000 km den fazla	42.597	94.9147
6	MY 2010 ile 2014	ABS	42.1564	93.9331
7	F 30000 ile 50000	Merkezi Kilit	41.4753	86.8907
8	F 30000 ile 50000	CD Çalar	41.3748	86.6803
9	F 30000 ile 50000	Elektrikli Ayna	40.5918	85.0399
10	MY 2010 ile 2014	Elektrikli Ön Cam	40.3142	89.8284
11	F 30000 ile 50000	Immobilizer	40.1314	84.0753
12	MY 2010 ile 2014	CD Çalar	39.9595	89.0378
13	MY 2010 ile 2014	MP3 Çalar	39.005	86.911
14	MY 2010 ile 2014	Merkezi Kilit	38.7416	86.3242
15	MY 2010 ile 2014	Elektrikli Ayna	38.5952	85.9981
16	MY 2010 ile 2014	Immobilizer	36.9222	82.2701
17	F 30000 ile 50000	MP3 Çalar	36.8497	77.2001
18	Dizel Araç	sdisi<=5	36.4063	76.6715
19	sdisi<=5	Dizel Araç	36.4063	81.5442
20	MY 2010 ile 2014	isofix	36.2187	80.7027
21	Dizel Araç	Hızlanma > 10 s	36.0439	75.9083
22	F 30000 ile 50000	Manuel Vites	36.0425	75.5091
23	F 30000 ile 50000	Kasetçalar	35.9563	75.3285
24	F 30000 ile 50000	Yükseklik > 1400 mm	35.7143	74.8214
25	Dizel Araç	Azamisurat<=200km/sa	35.6029	74.9795
26	Dizel Araç	Manuel Vites	35.4542	74.6663
27	MY 2010 ile 2014	Ayarlanabilir Dir.	35.2856	78.6234
28	MY 2010 ile 2014	Yükseklik > 1400 mm	34.6129	77.1247
29	F 30000 ile 50000	Hızlanma > 10 s	34.5935	72.4734
30	F 30000 ile 50000	Azamisurat<=200km/sa	34.5518	72.3861
31	MY 2010 ile 2014	Kasetçalar	34.5135	76.903
32	MY 2010 ile 2014	MTV>200	34.4504	76.7626
33	F 30000 ile 50000	sisFar	34.4107	72.0903
34	MY 2010 ile 2014	sisFar	33.9749	75.7029
35	F 30000 ile 50000	isofix	33.8356	70.8856
36	Dizel Araç	maxtorkb<=2500rpm	33.6878	70.9463
37	maxtorkb<=2500rpm	Dizel Araç	33.6878	94.9468
38	F 30000 ile 50000	Yakıt Depo <= 60 lt	33.5192	70.2227
39	MY 2010 ile 2014	Hızlanma > 10 s	32.7764	73.0326
40	MY 2010 ile 2014	Azamisurat<=200km/sa	32.6547	72.7613
41	MY 2010 ile 2014	Elektrikli Ön Cam	32.452	72.3096
42	MY 2010 ile 2014	Yakıt Depo <= 60 lt	32.1848	71.7143
43	MY 2010 ile 2014	Alaşımli Jant	31.6149	70.4446
44	MY 2005 ile 2010	20000 km den fazla	25.2097	99.1228
45	MY 2005 ile 2010	Merkezi Kilit	22.7233	89.3464
46	MY 2005 ile 2010	ABS	22.423	88.1656
47	Benzinli Araç	Merkezi Kilit	22.3055	89.182
48	MY 2005 ile 2010	Elektrikli Ön Cam	22.2743	87.5808
49	MY 2005 ile 2010	Immobilizer	22.1573	87.1207
50	F 20000 ile 30000	20000 km den fazla	21.7731	98.858
51	maxtorka>200	Dizel Araç	21.7215	92.8823
52	Benzinli Araç	20000 km den fazla	21.6362	86.5059
53	Benzinli Araç	Elektrikli Ön Cam	21.6248	86.4605
54	Benzinli Araç	ABS	21.5216	86.0476
55	MY 2005 ile 2010	CD Çalar	21.4704	84.4201
56	MY 2005 ile 2010	Elektrikli Ayna	20.6249	81.0955
57	Benzinli Araç	CD Çalar	19.8452	79.3451
58	Benzinli Araç	Immobilizer	19.7424	78.9341
59	Benzinli Araç	Elektrikli Ayna	19.606	78.3887
60	F 20000 ile 30000	Merkezi Kilit	19.3402	87.8119
61	F 20000 ile 30000	Elektrikli Ön Cam	19.0323	86.4139
62	MY 2005 ile 2010	Manuel Vites	18.8467	74.1037
63	F 20000 ile 30000	Manuel Vites	18.7676	85.2116

64	MY 2005 ile 2010	Kasetçalar	18.4147	72.4051
65	MY 2005 ile 2010	Yükseklik > 1400 mm	18.2238	71.6545
66	F 50000 ile 70000	ABS	18.21	94.6149
67	MY 2005 ile 2010	Sis Farı	17.9481	70.5705
68	Benzinli Araç	Sis Farı	17.8846	71.5062
69	Benzinli Araç	Kasetçalar	17.5862	70.3131
70	F 50000 ile 70000	Elektrikli Ayna	17.5767	91.3244
71	F 50000 ile 70000	CD Çalar	17.5705	91.2924
72	F 20000 ile 30000	ABS	17.3356	78.71
73	F 20000 ile 30000	Immobilizer	16.9372	76.9012
74	F 50000 ile 70000	Merkezi Kilit	16.8903	87.7581
75	F 50000 ile 70000	MP3 Çalar	16.8794	87.7015
76	F 50000 ile 70000	Elektrikli Ön Cam	16.7804	87.1871
77	F 50000 ile 70000	20000 km den fazla	16.5502	85.991
78	F 50000 ile 70000	Sis Farı	16.5095	85.7793
79	F 50000 ile 70000	ISOFIX	16.3702	85.0557
80	F 50000 ile 70000	Alaşımlı Jant	15.9287	82.7619
81	F 20000 ile 30000	Hidrolik Direksiyon	15.8705	72.0578
82	F 50000 ile 70000	Immobilizer	15.8454	82.3288
83	F 20000 ile 30000	Motor <= 100 beygir	15.843	71.9331
84	F 50000 ile 70000	Ayarlanabilir Dir.	15.7554	81.8611
85	MY 2000 ile 2005	20000 km den fazla	15.744	98.7961
86	F 50000 ile 70000	ESP	15.5811	80.9554
87	F 50000 ile 70000	Elektrikli Arka Cam	15.4669	80.3623
88	F 50000 ile 70000	Kasetçalar	15.1012	78.4623
89	F 50000 ile 70000	Yükseklik > 1400 mm	15.0941	78.4253
90	F 50000 ile 70000	MTV>200 TL	14.9108	77.4729
91	F 50000 ile 70000	Genişlik >1700 mm	14.3646	74.6351
92	F 50000 ile 70000	MY 2010 ile 2014	14.0373	72.9345
93	F 50000 ile 70000	Net Ağırlık > 1200 kg	13.9885	72.681
94	F 50000 ile 70000	Motor100-150 beygir	13.9625	72.5456
95	MY 2000 ile 2005	Merkezi Kilit	13.8573	86.9568
96	MY 2000 ile 2005	Elektrikli Ön Cam	13.5399	84.9652
97	MY 2000 ile 2005	ABS	12.6612	79.4513
98	MY 2000 ile 2005	Hidrolik Direksiyon	12.6608	79.4483
99	MY 2000 ile 2005	Manuel Vites	12.6044	79.0946
100	MY 2000 ile 2005	Immobilizer	11.495	72.133
MY : Model Yılı F : Fiyat MTV : Yıllık Motorlu Taşıtlı Vergisi Taksidi				

Tablo 4'deki kuralları şu şekilde özetlemek mümkündür. Dizel araçlar 20 000 km den fazla kullanılmıştır (2). Başka bir ifade ile dizel araçlar görece yüksek km değerleri ile satışa çıkmaktadır. Dizel araçlarda şehir dışı yakıt tüketimi 100 km ye 5 lt den azdır (18-19). Dizel araçlarda çoğunlukla 0-100 km hıza 10 saniyeden fazla sürede çıkmaktadır (21). Bu araçlarda çoğunlukla azami hız 200km/sa in altındadır (25) ve araçlar manuel vitese sahiptirler (26). Daha düşük devirlerde yüksek torka sahiptirler (36-37-51).

Fiyatı 30 000 ile 50 000 TL arasında olan araçlarda ABS (3), elektrikli ön cam (4), merkezi kilit (7), cd çalar (8), elektrikli ayna (9), mp3 çalar (17), manuel vites (22), kasetçalar (23), sis farı (33), isofix (35), immobilizer (11) donanımları bulunurken, bu araçlarda yükseklik 1400 mm nin üzerinde (24), 0-100 km/sa hıza çıkma süreleri 10 saniyeden fazla (29), yakıt depo hacmi 60 lt nin altındadır (38) ve azami hız 200km/sa in altındadır (30). Ayrıca 20 000 km den fazla kullanılmışlardır (1).

Model yılı 2010 ile 2014 olan araçlar 20000 km den fazla kullanılmıştır (5). Bu araçlarda ABS (6), elektrikli ön cam (10), CD çalar (12), MP3 çalar (13), merkezi kilit (14), elektrikli ayna (15), immobilizer (16), isofix (20), ayarlanabilir direksiyon (27), kasetçalar (31), sis farı (34), elektrikli ön cam (41), alaşımlı jant (43) donanımları bulunmaktadır. Ayrıca bu araçlarda yıllık Motorlu Taşıtlı Vergisinin bir taksidi 200 TL nin üzerindedir (32). Azami hız saatte 200 km nin altındadır (40) ve 0-100 km/sa hıza 10 saniyenin üzerinde bir zamanda çıkılabilmektedir (39). Yakıt depo hacmi 60 lt nin altındadır (42).

Model yılı 2005 ile 2010 arası olan araçlar 20 000 km den fazla kullanılmıştır (44). Merkezi kilit (45), ABS (46), elektrikli ön cam (48), immobilizer (49), CD çalar (55), elektrikli ayna (56), manuel vites (62), kasetçalar (64), sis farı (67) donanımları mevcuttur. Yükseklik 1400 mm nin üzerindedir.

Benzinli araçlarda genelde merkezi kilit (47), elektrikli ön cam (53), ABS (54), CD çalar (57), immobilizer (58), elektrikli ayna (59), sis farı (68), kasetçalar (69) donanımları bulunmaktadır. Genellikle 20 000 km den fazla kullanılmışlardır (52).



Fiyatı 20 000 TL ile 30 000 TL arasında olan araçlar genelde 20 000 km den fazla kullanılmıştır (50). Bu araçlarda merkezi kilit (60), elektrikli ön cam (61), manuel vites (63), ABS(72), immobilizer (73), hidrolik direksiyon (81) donanımları bulunmaktadır. Motor güçleri 100 beygirin altındadır (83).

Fiyatı 50 000 TL ile 70 000 TL arası olan araçlarda ABS (66), elektrikli ayna (70), CD çalar (71), merkezi kilit (74), MP3 çalar (75), elektrikli ön cam (76), sis farı (78), isofix (79), alaşımli jant (80), immobilizer (82), ayarlanabilir direksiyon (84), ESP (86), elektrikli arka cam (87), kasetçalar (88) donanımları bulunmaktadır. Bu araçlar 20 000 km den fazla kullanılmışlardır (77). Bu araçların yükseklikleri 1400 mm ni üzerindedir (89), genişlikleri 1700 mm nin üzerindedir (91), net ağırlıkları 1200 kg ın üzerindedir (93). Bu araçların yıllık motorlu taşıt vergisinin bir taksidi 200 TL nin üzerindedir (90). Motor güçleri 100 – 150 beygir arasındadır (94).

Son olarak model yılı 2000 ile 2005 olan araçlarda merkezi kilit (95), elektrikli ön cam (96), ABS(97), hidrolik direksiyon (98), manuel vites (99) ve immobilizer (100) donanımları bulunmaktadır. Bu araçlar da 20 000 km nin üzerinde kullanılmışlardır (85).



### 4.3. Sonuçların Görselleştirilmesi

Grafikler birçok gerçek hayat problemini başarılı bir şekilde modellemek için kullanılmaktadır (Balakishnan ve Ranganathan, 2012:1). Ağların görselleştirilmesi, sosyal ağlar, bilgi ağları, taşıma ağları gibi birçok ağ yapısının görselleştirilmesinde kullanılmaktadır (Cherven, 2013). Çalışmada elde edilen kuralların da şebeke (ağ) grafiği ile görselleştirilmesi mümkündür. Şekil 3'de kuralların şebeke grafiği yer almaktadır. Grafiğin merkezinde, kurallarda sıklıkla kullanılan değişkenler yer almaktadır. Grafiğin çevresinde ise daha az sayıda kuralda ismi geçen değişkenler yer almaktadır.

Eğer bir sınıftaki arkadaşlık ilişkilerini ağ grafiği yardımıyla görselleştirmek isteseydik, sonuçta ortaya çıkacak olan ağ grafiğinin merkezinde, popüler (çok fazla arkadaşı olan) öğrenciler yer alacaktı. Burada ise merkezde yer alan değişkenler çok fazla kuralda ismi geçen değişkenlerdir. Şekil 3 göz önünde bulundurulduğunda merkezi değişkenlerin model yılının 2010 ve 2014 ( $2010 < F \leq 2014$ ) arasında olması değişkenlerinin olduğu belirlenmektedir. Çevrede yer alan değişkenler ise daha az kuralda ismi geçen değişkenlerdir.

### Sonuç

Çalışmanın amacı, apriori algoritması yardımıyla ikinci el otomobil piyasasında satışa sunulan araçların genel özelliklerini belirlemektir. Apriori algoritması böyle bir amaç için ilk kez uygulanmaktadır. Böyle bir çalışma yapılmasının çeşitli avantajları vardır. İlk olarak ikinci el otomobil piyasası, veri madenciliği tekniği ile incelenmektedir. Bunun yanı sıra kullanılan değişkenlerden biri de fiyat olduğu için, sonuçta ortaya çıkan kuralların fiyatlandırma politikalarını şekillendirebilir nitelikte olduğunu söylemek mümkündür.

Çalışmanın amacını gerçekleştirebilmek için ilk olarak ikinci el otomobil piyasasından çok sayıda araca ilişkin bilgiler web kazıma tekniği ile bir araya getirilmiş ve bir veri tabanı oluşturulmuştur. Bu veri tabanında yer alan verilere, veri madenciliği yöntemlerinden biri olan apriori algoritması uygulanmıştır. Sonuçta ikinci el araç piyasasında işlem yapacak kişi ve kurumlar için, piyasayı daha iyi incelemelerine olanak sağlayacak bilgiler ortaya çıkmıştır.

Çalışmada, "dizel araçlar şehir dışında az yakar, dizel araçlar fazla hız yapamazlar, dizel araçların torku yüksektir, yeni araçların ve pahalı araçların vergisi yüksektir, pahalı araçlar daha geniş, daha ağırdır" gibi ifadeler birliktelik kuralları ile de doğrulanmıştır. Ayrıca, çalışmada oluşturulan kurallardan şu şekilde bir sonuç çıkarılabilir: fiyatı 30 000 TL ile 50 000 TL arasında (50 000 TL dahil) bir araç satın alınmak istenirse, araçlarda büyük olasılıkla ABS, elektrikli ön cam, merkezi kilit, CD çalar, elektrikli ayna ve manuel vites donanımları yer alacaktır. Aynı şekilde model yılı 2010 ile 2014 arasında (2014 dahil) yer alan ikinci el bir araç satın alınmak istendiğinde, araçta muhtemelen ABS, elektrikli ön cam, merkezi kilit, ayarlanabilir direksiyon gibi donanımlar yer alacak ve motorlu taşıtlar vergisi taksiti 200 TL nin üzerinde olacaktır. Diğer kurallara metin içinde değinilmiştir.

Çalışmanın bazı kısıtları da mevcuttur. Daha önce açıklanan nedenden ötürü sadece bir adet sitede yer alan ilanlar kullanılmıştır. Sitede kullanıcıların girdiği verilerin doğru olduğu kabul edilmektedir. Her ne kadar bazı bilgiler doğru girilmemiş olsa da, çalışmada kullanılan detaylı ön işleme yöntemi sayesinde veri setinde yer alan ve uç değer niteliğinde olan ilanların elenmesi sağlanmıştır.

Apriori algoritması ortaya çıktığından beri güçlendirilmiş algoritmalar da literatüre tanıtılmaktadır. İlerleyen çalışmalarda güçlendirilmiş apriori algoritmaları kullanılabilir. Çalışmanın karşılaşılan problemlerden biri de sayısal değişkenlerin kategorik değişkenlere çevrilmesi problemidir. Bazı sayısal değişkenlerde (vergi, uzunluk, yükseklik, net ağırlık vb) medyan değeri kullanılmışken, model yılı için 2000-2005 arası, 2005-2010 arası, 2010-2014 arası olarak belirlenmiştir. İlerleyen çalışmalarda sayısal değişkenler kategorik hale getirilirken farklı yöntemler kullanılabilir. Daha az ya da daha fazla sayıda kurallarla piyasa daha detaylı bir şekilde incelenebilir.

### Kaynakça

- Akçi, Y. (2016). İkinci El Otomobil: Tüketici Bakışıyla. *Adıyaman Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 8(22), 329:362.
- Asilkan, Ö., Irmak, S. (2009). İkinci El Otomobillerin Gelecekteki Fiyatlarının Yapay Sinir Ağları İle Tahmin Edilmesi. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(2), 375-391.
- Ay, D., Çil, İ. (2008). Migros Türk. A.Ş. de Birliktelik Kurallarının Yerleşim Düzeni Planlamada Kullanılması. *Endüstri Mühendisliği Dergisi YA/EM 2008 Özel Sayısı*, 21(2), 14-29.
- Aytaç, M.B., Bilge, H.Ş. (2013). Tele Pazarlama Verilerinin Birliktelik Kurallarıyla ve Crisp-Dm Yöntemiyle Analiz Edilmesi. *Aksaray Üniversitesi İİBF Dergisi*, 5(2),25-40.
- Balakrishnan, R., ve Ranganathan, K. (2012). *A Textbook of Graph Theory*. New York: Springer Science+Business Media.
- Cavique, L. (2007). A Scalable Algorithm for the Market Basket Analysis. *Journal of Retailing and Consumer Services*, 14, 400-407.
- Cherven, K. (2013). *Network Graph Analysis and Visualization with Gephi*. Birmingham: Packt Publishing.

- Daştan, H. (2016). Türkiye’de İkinci El Otomobil Fiyatlarını Etkileyen Faktörlerin Hedonik Fiyat Modeli İle Belirlenmesi. *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 18(1), 33-327.
- Doğan, B., Erol, B., Buldu, A. (2014). Sigortacılık Sektöründe Müşteri İlişkileri Yönetimi için Birlikte Kurallarının Kullanılması. *Marmara Fen Bilimleri Dergisi*, 3, 105-114.
- Doğrul, G., Akay, D., Kurt, M. (2015). Trafik Kazalarının Birlikte Kuralları ile Analizi. *Gazi Mühendislik Bilimleri Dergisi*, 1(2),265-284.
- Ecer, F. (2013). Türkiye’de 2. El Otomobil Fiyatlarını Tahmini ve Fiyat Belirleyicilerinin Tespiti. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 13(4),101-112.
- Erdem, S., Özdağoğlu, G. (2008). Ege Bölgesi’ndeki Bir Araştırma Ve Uygulama Hastanesinin Acil Hasta Verilerinin Veri Madenciliği İle Analiz Edilmesi. *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, 9(2),261-270.
- Erpolat, S. (2012). Otomobil Yetkili Servislerinde Birlikte Kurallarının Belirlenmesinde Apriori ve Fp-Growth Algoritmalarının Karşılaştırılması. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 12(1),37-146.
- Giudici, P. ve Figin, S. (2009). *Applied Data Mining for Business and Industry* (2nd ed). West Sussex: John Wiley & Sons Ltd.
- Han, J. ve Kamber, M. (2006). *Data Mining Concepts and Techniques* (2nd ed). San Fransisco: Elsevier.
- Hand, D., Mannila, H. ve Smyth, P. (2001). *Principles of Data Mining*. Cambridge: The MIT Press.
- Karabatak, M., İnce, M .C. (2004). Apriori Algoritması İle Öğrenci Başarısı Analizi. *Eleco Elektrik-Elektronik ve Bilgisayar Mühendisleri Sempozyumu*, Bursa.
- Kaur, M., Kang, S. (2016). Market Basket Analysis: Identify The Changing Trends of Market Data Using Association Rule Mining. *International Conference on Computational Modeling and Security, Procedia Computer Science*, 85, 78-85.
- Keleş, A.E., Kaya M. (2014). Duvar İnşa Edilmesinde Verimliliği Etkileyen Faktörlerin Apriori Veri Madenciliği Yöntemi Kullanılarak Analizi. *Akademik Bilişim Konferansları*, AB2014.
- Kurt P, M., Duru, N. (2015). Veri Madenciliği Teknikleri Kullanılarak Ortaokul Öğrencilerinin Sosyal Ağ Kullanım Analizi: Kocaeli İli Örneği. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 3, 508-517.
- Olson, D., ve Delen, D. (2008). *Advanced Data Mining Techniques*. Berlin: Springer.
- Özçakır, F.C., Çamurcu, Y. (2007). Birlikte Kuralı Yönetimi İçin Bir Veri Madenciliği Yazılımı Tasarımı ve Uygulaması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 6(12), 21-37.
- Pande, A., Abdel-Aty, M. (2009). Market Basket Analysis of Crash Data from Large Jurisdictions and Its Potential as a Decision Support Tool. *Safety Science*, 47, 145-154.
- Timor, M., Şimşek, U. T. (2008). Veri Madenciliğinde Sepet Analizi İle Tüketici Davranışı Modellemesi. *Yönetim*, 19,3-10.
- Witten, Ian. H., Frank, E. ve Hall, M.A. (2011). *Data Mining Practical Machine Learning Tool and Techniques* (3rd ed.). Burlington: Elsevier.

---

## Summary

*Automobiles are the second most popular assets that households want to possess. Moreover according to statistics published by TurkStat, number of automobiles that are handed over are increasing over the years. That is why this market needs a detailed examination by academicians to improve its effectiveness.*

*Various studies are performed to predict the price of second hand automobiles (Daştan, 2016; Asilkan ve Irmak, 2009; Ecer, 2003). Useful information about the predictors of second hand cars are emerged from these studies. However, dataset gathered from second hand car market have not been analyzed with association rules.*

*The purpose of this study is to make a descriptive study of second hand car market with association rules. Moreover, the rules are visualized with a network plot. Applying association rules to the dataset gathered from second hand car market can reveal the information of which two specialties are occurred together in most of the cars.*

*Data mining is the process of discovering patterns that are hidden in dataset (Witten et al. 2011:5). Data mining discovers that patterns by variety of algorithms. Three different algorithms are prominent in these algorithms namely Classification and Regression Tree (CART), backpropagation and apriori algorithm. By using apriori algorithm it is possible to discover boolean association rules in a dataset. The most popular application of apriori algorithm in business field is known as market basket analysis. Market basket analysis can directly affect the marketing strategies. Moreover, apriori algorithm is successfully applied for other purposes such as analyzing of student performance (Karatabak and Ince, 2004), customer relation management in insurance industry (Doğan et al, 2014), analyzing of emergency service dataset (Erdem and Özdağoğlu, 2008), analysing of accidents (Doğrul et al, 2015; Pande and Abdel-Aty, 2009).*

*In order to discover the rules, belong to the second-hand car market, a dataset is scraped from a major e-commerce website operating in Turkey. For this purpose, 211109 usable advertisements are collected. There are 73 variables are available for each advertisement. Analysis of the study is performed in MATLAB platform. Minimum support value is determined as 10% and minimum confidence parameter is fixed at 70%. By these parameters, the list of association rules are limited to a certain degree.*

*In this study, apriori algorithm is applied to a dataset scraped from second hand car market. At the end of the study usefully information is mined from the dataset which enable the figures of second hand car market to make efficient decisions. To the best knowledge of the author of this paper the apriori algorithm is applied first time for such a purpose. There are various benefits of making such application. First the market is analyzed with a data mining technique. Second, the results can shape the pricing policy in second hand market because one of the variables analyzed in this study is price. Rules such as “diesel cars use less fuel in out-of-town usage, diesel cars cannot go much faster, diesel cars have higher torque, expensive cars have higher weight and longer width” are discovered at the end of the study. There are some limitations of this study. First advertisements of a single website are used. Also, users are filling the information in advertisements. Users can fill the advertisements in a wrong manner. The information filled by users are accepted as true.*

---

