

Öğrencilerin Staj Verileri Üzerine Uygulanan Apriori Algoritması ile Birliktelik Kurallarının Çıkarılması ve Staj Eğiliminin Belirlenmesi

^{*1}Mehmet Taş, ²M. Fatih Adak, ²Nilüfer Yurtay

^{*1}Endüstri Mühendisliği Bölümü, Mühendislik Fakültesi, Sakarya Üniversitesi

²Bilgisayar Mühendisliği Bölümü, Bilgisayar ve Bilişim Bilimleri Fakültesi, Sakarya Üniversitesi

Özet

Bu çalışmada Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi Bilgisayar Mühendisliği Bölümü'nde öğrenim görmüş/gören öğrencilerin 1999-2012 yılları arasındaki zorunlu donanım ve yazılım stajı kayıtları kullanılarak bölüm öğrencilerinin staj eğilimleri belirlenmeye çalışılmıştır. Çalışmada veri madenciliği literatüründe kabul görmüş bir süreç olan CRISP-DM Metodolojisinden yararlanılmış; SPSS Clementine veri madenciliği programının Apriori algoritması aracı ile ilgili veriler üzerinde birliktelik kuralları analizi gerçekleştirilmiştir. Elde edilen bulgulardan, bölüm staj sorumlularına, staj yapan bölüm öğrencileri ile ilgili bilgiler verilmeye çalışılarak staj etkinliğini arttırmaya ve bölüm staj politikasını iyileştirmeye yönelik kararlar almalarında destek sağlanmıştır.

Anahtar kelimeler: veri madenciliği, birliktelik kuralları, apriori algoritması

Abstract

In this study, hardware and software internships data of Computer Engineering students in Sakarya University between the years 1999-2012, have been used to determine students trends and achievements in internships. The Apriori algorithm that accepted in data mining literature was used by SPSS Clementine tool which is from CRISP-DM methodology. As a result of Apriori algorithm association rules have been obtained. The obtained results could provide efficiency in internships to whose responsible from internships at the department.

Key words: data mining, association rules, apriori algorithm

1. Giriş

Her geçen gün artmaya devam eden ve veri tabanlarında saklanan dijital veriler, bu verilerden etkin ve verimli şekilde faydalanma isteğini de beraberinde getirmektedir. Bu istek, “Veri Tabanlarında Bilgi Keşfi (VTBK)” başlığı altında toplanan çalışmalara sebep olmuştur. VTBK farklı kısımlardan oluşan bir süreçtir. “Veri Madenciliği” bu sürecin model kurma ve değerlendirme aşamalarını kapsamaktadır (Erpolat, 2012). Basit bir tanım yapmak gerekirse, veri madenciliği, büyük ölçekli veriler arasından değeri olan bir bilgiyi elde etme işidir. Bu sayede veriler arasındaki ilişkileri ortaya koymak ve gerektiğinde ileriye yönelik tahminlerde bulunmak mümkün görülmektedir (Özkan, 2008).

*Corresponding author: Adres: Endüstri Mühendisliği Bölümü, Mühendislik Fakültesi, Sakarya Üniversitesi, 54187, Sakarya TÜRKİYE. E-mail adres: tas@sakarya.edu.tr, Phone: +902642955463

Mail Adr: M. Tas (tas@sakarya.edu.tr), M. F. Adak (fatihadak@sakarya.edu.tr), N. Yurtay (nyurtay@sakarya.edu.tr)

Bir arada gerçekleşen olayları analiz etmek veri madenciliğinin kapsamı içine girmektedir. Olayların birlikte gerçekleşme durumlarını çözümleyen veri madenciliği yöntemlerine birliktelik kuralları adı verilmektedir (Özkan, 2008). Birliktelik kuralları bir arada sık olarak görülen ilişkilerin ortaya çıkarılmasını ve özetlenmesini sağlamaktadır.

Veri madenciliğinin en önemli fonksiyonlarından birisi kural çıkarmadır. Geçmiş verilerin yakın gelecekte de geçerli olacağı varsayımı ile hareket ettiğimizde, geçmiş verilerden elde edilen kurallar gelecek için doğru yapmamızı kolaylaştıracaktır (Alpaydın, 2000). Kural çıkarmak amacıyla geliştirilmiş farklı yöntemler bulunmaktadır. Apriori algoritması bunlardan biri olup, veri madenciliğinde sık tekrarlanan öğeleri keşfetmek için kullanılan en temel yöntemdir.

Apriori Algoritmasının kullanıldığı bu çalışmada Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi Bilgisayar Mühendisliği Bölümü öğrencilerine ait zorunlu donanım ve yazılım stajı verileri kullanılarak öğrencileri staj eğilimleri belirlenmeye çalışılmıştır. Elde edilen bulgulardan, bölüm staj sorumlularına, staj yapan bölüm öğrencileri ile ilgili bilgiler verilmeye çalışılarak staj etkinliğini arttırmaya ve bölüm staj politikasını iyileştirmeye yönelik kararlar almalarında destek sağlanmıştır.

1.1. Önceki çalışmalar

Veri madenciliğinin günümüzde yaygın bir kullanım alanı bulunmaktadır. Ülkemizde eğitim alanında birliktelik kuralı algoritmaları uygulanarak gerçekleştirilmiş veri madenciliği çalışmaları incelendiğinde, yapılan bir çalışma üniversitede okumakta olan lisans öğrencilerinin eğitim amacıyla internet kullanımlarının içerdiği gizli veya açık örüntüleri, veri madenciliği yöntemleriyle tespit etmeye yöneliktir. Kaynak teşkil eden verilere üç adet veri madenciliği algoritması (karar ağaçları, kümeleme ve birliktelik kuralları) uygulanmış öğrencilerin eğitimsel amaçlar için internet kullanmalarına yönelik davranış ve düşüncelerinde sık görülen örüntüler bulunmuştur (Bozkır, vd., 2008). Yapılan diğer bir çalışmada apriori algoritması kullanarak, öğrencilerin başarı durumlarını tespit etmek amacıyla öncelikle, Apriori algoritmasını kullanan bir Matlab programı hazırlanmıştır. Sonra bu program, üniversite öğrencilerinin birinci ve ikinci sınıfta aldıkları genel kültür derslerine ait notlara uygulanarak bu derslerdeki başarı durumları tespit edilmiş ve buradan elde edilen kurallar irdelenmiştir (Karabatak ve İnce, 2004). Başka bir çalışmada Selçuk Üniversitesi'nde kullanılan Öğrenci İşleri Otomasyonundan elde edilen veriler üzerinden gelecekle ilgili plan yapılabilmesi için anlamlı bilgiler çıkarılmıştır. Bu bilgiler sayesinde üniversiteyi yeni kazanan bir öğrencinin, üniversitedeki başarısına etki eden etmenler göz önüne çıkarılmıştır (İnan, 2003). Diğer bir çalışmada ise ÖSS verilerine karar ağaçları ve birliktelik kuralları yöntemleri uygulanarak öğrencilerin tercih profilleri ortaya konulmaya çalışılmıştır (Dolgun, vd., 2007). Bir başka çalışmada öğrencilerin okullarındaki rehberlik servisinden aldıkları hizmetlerin, not ortalamaları, düzenli çalışma alışkanlıkları, aile durumu gibi bilgiler dikkate alınarak nasıl bir dağılım gösterdiği, verilen hizmetlerin diğer kriterler arasında nasıl ilişkiler olduğu ve bu kriterler arasında ne gibi kuralların bulunduğu tespit edilmeye çalışılmıştır (Sönmez, 2012).

1.2. Veri Madenciliği

Veri madenciliği, büyük hacimli veri yığınları içerisinde karar alabilmek için potansiyel olarak faydalı olabilecek, uygulanabilir ve anlamlı bilgilerin çıkarılmasına verilen addır (Argüden ve Erşahin, 2008). Veri madenciliği döngüsü, veri yığınlarını elden geçirmekle başlayarak, analiz sonucunda ortaya çıkan sonuçların uzman gözüyle yorumlanması ile tamamlanır.

Veri madenciliği modelleri, temelde iki ana başlıkta incelenmektedir. Birincisi, elde edilen örüntülerden sonuçları bilinmeyen verilerin tahmini için kullanılan tahmin edici modeller, diğeri ise eldeki verinin tanımlanmasını sağlayan tanımlayıcı modellerdir. Veri madenciliğinde kullanılan yöntemler kümeleme analizi, birliktelik kuralları ve ardışık zamanlı örüntüler, sınıflandırma ve regresyon analizi ve zaman serileri analizi olarak belirtilebilir (Argüden ve Erşahin, 2008).

1.3. Birliktelik Kuralları

Birliktelik kuralları veri madenciliğinin tanımlayıcı modellerindendir. Birliktelik kuralları büyük veri kümeleri içinde farklı veriler arasındaki birliktelik ilişkilerini bulma işlemidir. Birliktelik analizi, belirli bir veri kümesinde yüksek sıklıkta birlikte görülen özellik değerlerine ait ilişki kurallarının keşfidir (Argüden ve Erşahin, 2008).

Birliktelik kurallarının en önemli uygulama alanları arasında Pazar Sepet Analizi, çapraz-pazarlama, promosyon analizleri, katalog ve yerleşim düzeni tasarımları bulunmaktadır (Erpolat, 2012).

1.3. Apriori Algoritması

Birliktelik kuralı çıkarımında kullanılan Apriori Algoritması, “prior (önceki)” kelimesinden türetilmiştir ve bilgileri bir önceki adımdan almaktadır. Tekrarlayan bir niteliğe sahiptir ve veri tabanlarında sık geçen veri kümelerinin keşfedilmesinde kullanılmaktadır. Sık geçen veri kümelerini bulmak için birçok kez veri tabanını taramak gerekmektedir. İlk taramada bir elemanlı minimum destek ölçütünü sağlayan sık geçen veri kümeleri bulunmakta, izleyen taramalarda bir önceki taramada bulunan sık geçen veri kümeleri aday kümeler adı verilen yeni potansiyel sık geçen veri kümelerini üretmek için kullanılmaktadır. Aday kümelerin destek değerleri, tarama sırasında hesaplanmakta ve aday kümelerden minimum destek ölçütü sağlayan kümeler o geçişte üretilen sık geçen veri kümeleri olmaktadır. Sık geçen veri kümeleri bir sonraki geçiş için aday küme olmaktadır. Bu süreç yeni bir sık geçen veri kümesi bulunamayana kadar devam etmektedir (Gülce, 2010).

Pazar-sepet çözümlemelerinde satılan ürünler arasındaki ilişkileri ortaya koymak için “destek” ve “güven” gibi iki ölçütten yararlanılır. ‘Kural destek ölçütü’ bir ilişkinin tüm alışverişler içinde hangi oranda tekrarlandığını belirler. ‘Kural güven ölçütü’, A ürün grubunu alan müşterilerin B ürün grubunu da alma olasılığını ortaya koyar. A ürün grubunu alanların B ürün grubunu da alma durumu, yani birliktelik kuralı $A \rightarrow B$ biçiminde gösterilir. Bu durumda kural destek ölçütü

Denklem 1’deki gibi ifade edilir (Özkan, 2008).

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

Denklem 1’de $\text{sayı}(A,B)$ destek sayısı A ve B ürün gruplarını birlikte içeren alışveriş sayısını göstermektedir. N ise tüm alışverişlerin sayısını göstermektedir. A ve B ürün gruplarının birlikte satın alınması olasılığını ifade eden kural güven ölçütü ise Denklem 2’deki gibi hesaplanır (Özkan, 2008).

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

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

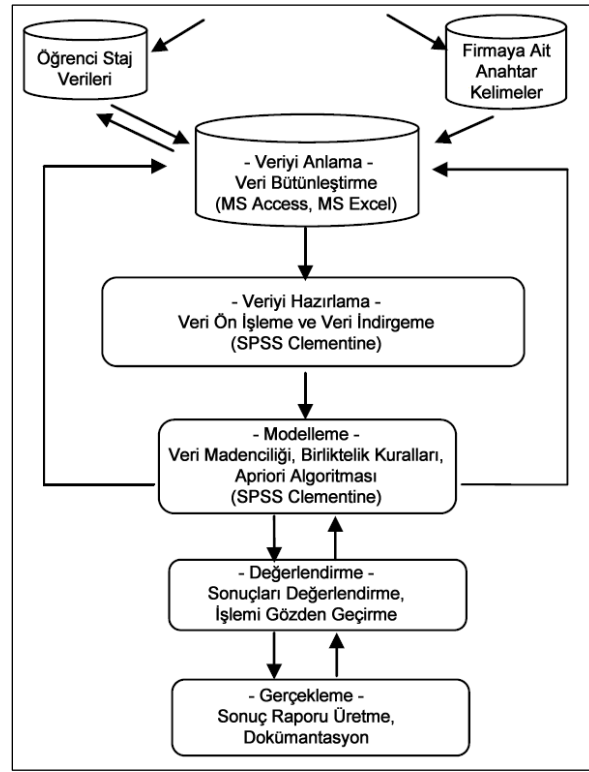
2. Metodoloji

Bu çalışmada Daimler Chrysler, SPSS, NCR ve OHRA tarafından geliştirilen ve veri madenciliği literatüründe kabul görmüş bir süreç olan CRISP-DM metodolojisinden yararlanılmıştır. CRISP-DM metodolojisine göre veri madenciliği süreci altı safhadan oluşmaktadır: İşin tanımlanması, verinin tanımlanması, verinin hazırlanması, modelleme, değerlendirme, uygulama (Chapman, vd., 2000).

2.1. İşin tanımlanması

Bu çalışmanın amacı Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi Bilgisayar Mühendisliği Bölümü öğrencilerine ait zorunlu donanım ve yazılım stajı verilerini kullanarak öğrencilerin staj eğilimlerini belirlemeye çalışmak ve bölüm staj sorumlularına, staj yapan bölüm öğrencileri ile ilgili bilgiler vermektir. Böylece yetkililerin, staj etkinliğini arttırmaya ve bölüm staj politikasını iyileştirmeye yönelik kararlar almalarında destek sağlanacaktır. Şekil 1’de çalışmanın aşamaları gösterilmiştir.

Bu çalışmada SPSS şirketinin veri madenciliği için geliştirdiği SPSS Clementine programının Apriori algoritması ve veri tabanı olarak Microsoft Access ve Excel kullanılmıştır.



Şekil 1. Çalışmanın kapsamı

2.2. Verinin tanımlanması

Bu aşamada, başlangıç verileri toplanmış, tek bir tablo haline getirilmiş; veriler tanımlanmış, genel olarak incelenmiş ve veri kalitesi doğrulanmıştır. Çalışmada Sakarya Üniversitesi Bilgisayar Mühendisliği Bölümü'nde öğrenim görmüş/gören öğrencilerin 1999-2012 yılları arasındaki zorunlu donanım ve yazılım stajı kayıtları kullanılmıştır. Çalışmada kullanılacak başlangıç verileri, adı geçen Bilgisayar Mühendisliği Bölümü'nün staj takip otomasyonundan alınmıştır. Edinilen öğrenci staj verileri Access veri tabanı dosyası şeklindedir. Staj veri tabanındaki alanlar sırasıyla sıra no, öğrenci no, staj çeşidi, staj süresi, onay süresi, onay tarihi, staj başlangıç tarihi, staj bitiş tarihi, staj yapılan firmanın adı, firma adresi, firmanın bulunduğu şehir, firmanın telefon numarası ve staj komisyonu kodudur.

İkinci olarak, bölümden alınan staj verileri dosyasındaki firma bilgilerinden hareketle her firma tek tek internet üzerinden araştırılmış, firma hakkında daha detaylı bilgi edinilmesini sağlayacak firmaya özel maksimum beş kelime kayıt edilmiştir. Bu anahtar kelimeler seçilirken firmanın faaliyet gösterdiği sektör, hizmetleri, çözümleri, vb. detaylar dikkate alınmıştır. Örneğin, KoçSistem Bilgi ve İletişim Hizmetleri A.Ş.'de için anahtar kelimeler “bilişim teknolojileri”, “sunucu sistemleri”, “sistem kurulum ve destek”, “danışmanlık” ve “bulut çözümleri” olarak seçilmiştir. Diğer örnekte, Sakarya Üniversitesi Bilgi İşlem Daire Başkanlığı için anahtar kelimeler “kamu kurumu”, “üniversite” ve “bilgi işlem” olarak belirlenmiştir. Öğrenci staj verileri ve firmaya ait anahtar kelimeler Excel ortamına alınmış ve burada incelemeye tabi tutulmuştur.

Oluşturulan analiz dosyasında 1113 öğrencinin zorunlu donanım ve yazılım stajından oluşan 2116 satır kayıt yer almaktadır. Bu öğrencilerin 814'ü I. öğretim, 239'u II. öğretim ve 60'ı karma öğretime kayıtlıdır. Firma adı sütunu incelendiğinde 853 farklı staj yapılan firma olduğu görülmektedir. Bu firmalar 68 farklı şehirde konuşlanmıştır ve bu şehirlerin beşi yurtdışında bulunmaktadır.

2.3. Verinin hazırlanması

Bu basamakta modelleme yapılabilmesi için gerekli olan veri hazırlığı (seçme, temizleme, yapılandırma, birleştirme, biçimlendirme) işlemleri yapılmıştır. Verilerin Excel ortamında toplanmasının ardından staj yapılan şehir sütunu ilçe ve şehir olmak üzere ikiye sütuna dönüştürülmüş, şehir sütununa yazılan bu iki değer ilgili sütunlarına dağıtılmıştır. Şehir kaydı bulunmayan firmalar için firmanın iletişim bilgilerine ulaşarak şehir bilgileri tamamlanmıştır. Farklı içerikte ve biçimde girilmiş firma ve şehir isimleri tespit edilerek tek tipe indirgenmiştir. Analizde gerek duyulmayacağı düşünülerek firma adres, ilçe ve bilgilerini içeren sütunlar ile staj komisyonu kodu sütunu elenmiştir. Öğrenci numaraları sütunu kullanılarak öğrencilerin üniversiteye giriş yılları ve hangi öğretim türüne kayıtlı olduğu hesaplanmış, yeni açılan giriş yılı ve öğretim türü sütunlarına eklenmiştir. Benzer şekilde staj başlangıç ve bitiş tarihlerinin bulunduğu sütunlar kullanılarak stajın yapıldığı yıl hesaplanmış, yeni açılan staj yılı sütununa eklenmiştir. Ayrıca staj süresi ve onay süresi sütunları kullanılarak staja ait kabul veya ret durumu hesaplanmış, yeni açılan staj onayı sütununa eklenmiştir. Staj onayı sütunu incelendiğinde stajı reddedilen öğrencilerin yüzdesinin 1,8 olduğu görülmüştür. Bu bakımdan onay durumu alanı ve ek olarak, amaca katkı sağlamayacağı düşüncesiyle öğrenci no, staj süresi, onay süresi, onay tarihi, staj başlangıç tarihi ve staj bitiş tarihi alanları analize dahil edilmemesine karar verilmiştir. Staj yapılan 853 farklı kurumdan sadece %3'ünde 10 ve üzeri defa, %70'inde ise yalnız 1 kez staj yapıldığının tespit edilmesi üzerinde gereksiz işlem yükü meydana getirmemek için en az 10 staj yapılan yerlerin firma adı sütunu boşaltılmış, diğer sütunları aynen bırakılmıştır. Son olarak her firma için ayrı ayrı girilen maksimum beş anahtar kelime için beş sütun eklenmiş ve internet üzerinden edinilen firma detayları kullanılarak ilgili sütunlara ilgili veriler aktarılmıştır. SPSS Clementine'de Apriori algoritmasını çalıştırabilmek için verilerin tabular formata getirilmesi gerekmektedir. Tabular formattaki kayıtlar için sütunlarda giriş yılı, öğretim türü, staj çeşidi, staj yılı, firma adı, firma türü (beş adet) ve firmanın yer aldığı şehir yer alırken, satırlarda öğrencilere ait staj kayıtları yer almaktadır. Analiz için kullanılacak veriler Excel ortamında hazırlandıktan sonra SPSS Clementine'de Excel dosyasına bağlanma işlemleri yapılarak veri kaynağına erişim sağlanmıştır.

2.3. Modelleme

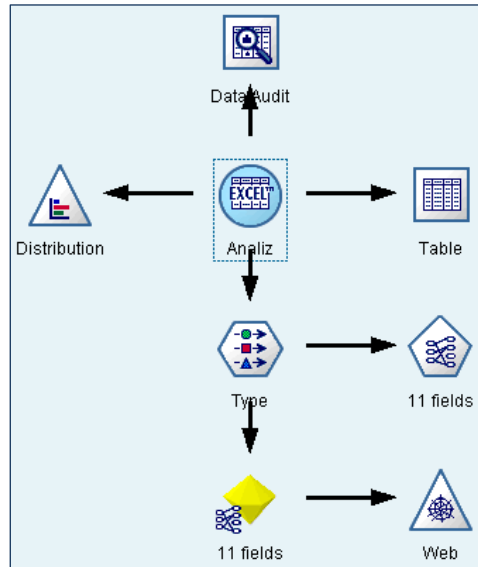
Clementine'da veri kaynağına erişim sağlandıktan sonra model oluşturulmuştur. Modelde yapılan işlemler;

- Data Audit düğümüyle veri kalitesi kontrol edilmiştir.
- Table düğümü ile veri kaynağından çekilen veriler kontrol edilmiştir.
- Distribution düğümü ile verilerin dağılımları ve sıklıkları hakkında bilgi alınmıştır. Distribution düğümünden elde edilen istatistiksel bilgiler Tablo 1'de örneklendiği gibidir.

Tablo 1. Tanımlayıcı istatistikler

Bilgi	İçerik	Sıklık	Yüzde
En çok staj yapılan beş kurum	SAÜ Bilgisayar Müh. Bölümü	164	7,75
	Türk Telekom İl Müdürlükleri	109	5,15
	Ford Otosan A.Ş.	97	4,58
	SAÜ Bilgi İşlem Daire Bşk.	77	3,64
	Türkiye Vagon Sanayi A.Ş.	53	2,5
En çok staj yapılan beş şehir	İstanbul	750	35,44
	Sakarya	661	31,24
	Kocaeli	213	10,07
	Ankara	116	5,48
	Bursa	48	2,27
En çok kullanılan firma anahtar kelimeleri	Kamu Kurumu	515	11,18
	Telekomünikasyon	515	11,18
	İmalat	442	9,59
	Üniversite	334	7,25
	Yazılım	299	6,49

Yapılan bu çalışmada birliktelik kuralları algoritmalarından bilinen en temel algoritma olan Apriori algoritmasından yararlanılmıştır. Yapılan çalışmalar sonucunda Apriori algoritmasında, minimum destek değeri %5 ve güven değeri %20 alınmıştır. Şekil 2’de Clementine’de kurulan modelin son aşamasındaki görüntüsü verilmiştir.

**Şekil 2.** Clementine’de kurulan modelin görüntüsü

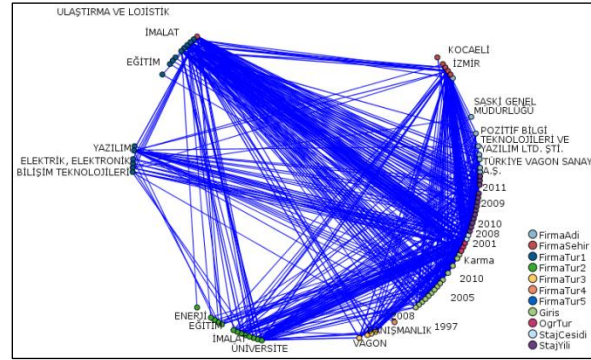
3. Bulgular

Sonuç olarak toplamda 989 tane kural oluşmuştur. Modelin çalıştırılması sonucunda elde edilen kurallardan bazıları Tablo 2’de verilmiştir. Derece ile ifade edilen değer 1’den ne kadar büyük olursa kuralın kapsamı artmaktadır. Analiz sonucunda oluşan birliktelik kurallarını anlamlandırmak gerekirse, örneğin Tablo 2’deki ilk kuraldan “Yazılım stajını Sakarya’daki bir kamu kurumunda yapan öğrenciler %88,74 olasılıkla Üniversite’de staj yapmaktadır. Bu durumların veri setindeki işlemlerde birlikte bulunma olasılığı %7,14’tür.” manası çıkmaktadır. Oluşturulan veri setinden öğrencilerin direkt olarak stajlarıyla ilgili birliktelikler çıktığı gibi, staj yapılan firmalara ait sonradan girilmiş anahtar kelimelerin yardımıyla, sondan bir önceki kuralda belirtildiği üzere “Staj yapılan firmalardan bilişim teknolojileri alanında faaliyet gösteren firmalar %25,47 olasılıkla sunucu sistemlerine yönelik hizmet vermektedir.” şeklinde endüstriyel anlamlar da çıkarılabilmektedir.

Veri setinde analize dahil edilmiş alanlar arasındaki ilişkilerin kuvvetini görebilmek için Clementine’de web grafiğinden yararlanılmıştır. Web grafiğindeki çizgi kalınlığı ilişkinin baskınlığını göstermektedir. Çizgi kalınlığı arttıkça ilişkinin kuvvetliliği de artmaktadır. Şekil 3’de gösterilen grafikte çizgi değeri %20 olarak ayarlanmıştır. Bu grafikte aralarındaki ilişki %20’den büyük ve eşit olan içerikler yer almaktadır.

Tablo 2. Birliktelik kuralları

Öncül	Sonuç	Destek	Güven	Derece
Staj Türü: Yazılım Şehir: Sakarya Firma Türü: Kamu Kurumu	Firma Türü: Üniversite	7,14	88,74	6,02
Firma Türü: Bilişim Teknolojileri	Şehir: İstanbul	5,01	85,85	2,42
Firma Türü: İmalat Şehir: Kocaeli	Firma Adı: Ford Otosan	6,38	68,15	14,87
Firma Türü: Bilgisayar Sarf Malzemeleri Staj Türü: Donanım	Firma Türü: Teknik Destek	5,48	62,07	10,42
Firma Türü: İmalat Staj Çeşidi: Yazılım	Firma Türü: Otomotiv	7,47	51,27	6,46
Firma Türü: Üniversite Staj Türü: Donanım	Firma Adı: SAÜ Bilgisayar Müh.	7,56	41,88	5,40
Staj Türü: Yazılım Şehir: İstanbul	Firma Türü: Yazılım	17,49	28,38	2,06
Şehir: Sakarya Firma Türü: Üniversite	Firma Türü: Bilgi İşlem	12,85	28,31	5,93
Firma Türü: Bilişim Teknolojileri Staj Yılı: 2010 I. Öğretim	Firma Türü: Sunucu Sistemleri Firma Türü: İmalat	5,01 5,86	25,47 22,58	12,53 1,25



Şekil 3. Analiz verileri arasındaki ilişkilerin grafiği

Sonuçlar

Birliktelik kurallarından elde edilen kurallar sonucunda bölüm staj sorumlularına, staj yapan bölüm öğrencileri ile ilgili bilgiler verilmeye çalışılarak staj etkinliğini arttırmaya ve bölüm staj politikasını iyileştirmeye yönelik kararlar almalarında destek sağlanmıştır. Öğrenci staj verilerdeki alanların iyileştirilmesi ve staj yapılan kurumlara ait verilerin detaylandırılması ile analizin karar vericilere sağladığı bilgi desteği maksimize edilebilecektir. Birliktelik kurallarından elde edilen sonuçlara göre öğrencilerin staj olarak en çok yazılım ile ilgili firmaları tercih ettiği geri kalanların ise çoğunlukla kendi okudukları üniversitede staj yaptıkları gözlemlenmiştir.

Referanslar

- Alpaydın, E. (2000). Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri. Bilişim 2000 Veri Madenciliği Eğitim Semineri, İstanbul.
- Argüden, Y., Erşahin, B. (2008). Veri Madenciliği: Veriden Bilgiye, Masraftan Değere. 1. Basım, ARGE Danışmanlık Yayınları, İstanbul.
- Bozkır, A.S., Gök, B., Sezer, E. (2008). Üniversite Öğrencilerinin İnterneti Eğitimsel Amaçlar İçin Kullanmalarını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti. Bilimde Modern Yöntemler Sempozyumu BMYS'2008, Eskişehir, 833–842.
- Chapman, P. (2000). CRISP-DM 1.0 Step-by-step data mining guide. SPSS Inc.
- Dolgun, M. Ö., Özdemir, T.G., Deliloğlu, S.. (2007). Öğrenci Seçme Sınavında (ÖSS) Öğrencilerin Tercih Profillerinin Veri Madenciliği Yöntemleriyle Tespiti, Bilişim 07 Kongresi, Ankara.
- Erpolat, S. (2012). Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması. AÜ Sosyal Bilimler Dergisi. C:12, S:1, ss. 151-166.
- Gülce, A.C. (2010). Veri Madenciliğinde Apriori Algoritması ve Apriori Algoritmasının Farklı Veri Kümelerinde Uygulanması. Trakya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği A.B.D. Yüksek Lisans Tezi.
- İnan, O. (2003). Veri Madenciliği, Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü.

Karabatak, M., İnce, M.C. (2004). Apriori Algoritması İle Öğrenci Başarısı Analizi. Eleco'2004 Elektrik-Elektronik ve Bilgisayar Mühendisleri Sempozyumu, Bursa.

Özkan, Y. (2008). Veri Madenciliği Yöntemleri, 1. Basım, Papatya Yayınları.

Sönmez, Y. (2012). Masaüstü Rehberlik Otomasyonu Gerçekleştirimi ve Verilerin Birliktelik Kuralları İle Değerlendirilmesi. Fırat Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Bilgisayar Eğitimi A.B.D. Yüksek Lisans Tezi.