

Apriori Algoritması ile Teknik Seçmeli Ders Seçim Analizi

Emre Güngör^{1,2}, Nesibe Yalçın^{1,2}, Nilüfer Yurtay³

¹ Bilecik Şeyh Edebali Üniversitesi, Bilgisayar Mühendisliği Bölümü, 11210, Merkez, Bilecik

² Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar ve Bilişim Mühendisliği A.B.D., 54187, Serdivan, Sakarya

³ Sakarya Üniversitesi, Bilgisayar Mühendisliği Bölümü, 54187, Serdivan, Sakarya

emre.gungor@bilecik.edu.tr, nesibe.yalcin@bilecik.edu.tr, nyurtay@sakarya.edu.tr

Özet: Bu çalışmada, Veri Madenciliği'nde Birliktelik Kuralı çıkarım algoritmalarından biri olan Apriori algoritması kullanılarak Bilecik Şeyh Edebali Üniversitesi Bilgisayar Mühendisliği Bölümü öğrencilerinin teknik seçmeli ders seçimlerindeki nedenlerin belirlenmesi amaçlanmaktadır. Bu amaçla, 2., 3. ve 4. sınıf öğrencilerinden anket yöntemi ile gerekli bilgiler alınmıştır. Daha sonra bu bilgiler üzerinde gerekli dönüşümler yapılarak Apriori Algoritması için uygun hale getirilmiştir. Elde edilen veriler Matlab® programlama dili ile gerçekleştirilen yazılıma gönderilmiş ve otomatik olarak birliktelik kurallarının çıkarılması sağlanmıştır. Bu sayede, öğrencilerin teknik seçmeli ders seçimlerinde hangi kriterleri göz önünde bulundurdıkları belirlenmiştir.

Anahtar Sözcükler: Apriori Algoritması, Birliktelik Kuralı, Veri Madenciliği.

Selection Behavior Analysis of Technical Elective Courses Using Apriori Algorithm

Abstract: In this study, it has been aimed to determine Bilecik Şeyh Edebali University, Department of Computer Engineering students' selection reasons of technical elective courses using Apriori Algorithm which is one of the Association Rule mining algorithms in Data Mining. For this purpose, required information was taken from 2., 3. and 4. grade students by questionnaire. Then the information was transformed in a way that is suitable for Apriori algorithm. The obtained data was sent to a software that was developed via Matlab® Software Package, and association rules were extracted automatically using the software. Thus, which criteria students consider in selections of technical elective courses was found.

Keywords: Apriori Algorithm, Association Rules, Data Mining.

1. Giriş

1990'lerde ortaya çıkan veri madenciliği, veri ambarlarındaki gizlenmiş potansiyel olarak faydalı bilgileri ortaya çıkarma, daha sonra bu bilgileri karar verme ve uygulama aşamasında kullanma sürecidir. Veri Madenciliği çözüm için gerekli bilgileri sağlamakta ancak kendi başına bir çözüm üretmemektedir [1, 2].

Veri madenciliği teknikleri, tanımlayıcı ve tahmin edici teknikler olmak üzere ikiye ayrılır: Tanımlayıcı teknikler, karar vermeye yardım edecek verilerin tanımlanmasını sağlar. Birliktelik kuralları (association rules)

ve kümeleme (clustering) tanımlayıcı tekniklere örnek olarak verilebilir. Bunlar Tahmin edici teknikler ise sonuçları bilinen verileri kullanarak sonuçları bilinmeyen veri kümelerinin sonuçlarının tahmin edilmesini sağlar. Gerileme (regression), sınıflandırma (classification) ve sapma (deviation) tahmin edici tekniklerdendir [2].

Bu çalışmada, gerçek veriler üzerinden tanımlayıcı tekniklerden biri olan birliktelik kurallarının çıkarılması amaçlanmıştır. Bu çalışmanın ikinci bölümünde, birliktelik kuralları ve birliktelik kuralı çıkarım algoritmalarından biri olan Apriori

algoritması üzerinde durulmuştur. Apriori algoritması incelenmiş, algoritma adımları ve parametreleri verilmiştir. Üçüncü bölümde uygulama detaylandırılmıştır. Son olarak dördüncü bölümde uygulamanın sonuçlarına yer verilmiştir.

2. Birliktelik Kuralları

Birliktelik kuralları, büyük veri kümeleri arasındaki birliktelik ilişkilerini bulurlar. Kullanışlılığı ve kolay anlaşılması gibi nedenlerden ötürü ekonomi, eğitim, telekomünikasyon, e-ticaret ve pazarlama gibi birçok alanda geniş bir kullanım alanına sahiptir [3, 4].

Birliktelik kurallarının en yaygın kullanıldığı alan süpermarket uygulamalarıdır. Bu uygulamalar, literatürde market sepeti analizi (market basket analysis) olarak geçmektedir. Market sepeti analizi, hangi ürünleri bir arada aldıkları bilgisini ortaya çıkararak müşteri alışkanlıkları analizlerinde, promosyon çalışmalarında, mağazanın düzenlenmesinde, stokların kontrolünde, etkili satış stratejilerinin geliştirilmesinde fayda sağlamaktadır [3].

2.1 Apriori Algoritması

Apriori algoritması, Agrawal ve Srikant [5] tarafından 1994 yılında geliştirilmiştir. Bu algoritma en yaygın kullanılan ve bilinen birliktelik kuralı çıkarım algoritmasıdır. Algoritmanın ismi, sık geçen öge kümelerinin önsel bilgisini kullanmasından, diğer bir ifadeyle bilgileri bir önceki adımdan almasından bir önceki (prior) anlamına gelen “apriori” dir [4].

Apriori algoritmasının temel yaklaşımı, “Eğer k-öge kümesi minimum destek kriterini sağlıyorsa, bu kümenin alt kümeleri de minimum destek kriterini sağlar.” şeklindedir. Öge küme, 1 veya daha fazla elemandan oluşan kümedir. k-öge küme (k-itemset) ise içinde k adet öge bulunan kümedir [3].

Apriori algoritması, (k+1) sık geçen öge kümesini bulmak için k sık geçen öge kümesine ihtiyaç duyar [6, 7]. Sık geçen öge kümeleri bulmak için ilk olarak minimum destek kriterini sağlayan 1-sık geçen öge kümeleri bulunur, 2- sık geçen öge kümeleri bulmak için 1- sık geçen öge kümeleri, 3- sık geçen öge kümeleri bulmak için 2- sık geçen öge kümeleri kullanılır. Bu süreç, algoritma sık geçen öge kümeleri bulamayınca kadar devam eder [6].

2.2 Apriori Algoritmasının Parametreleri

Apriori algoritmasının çalışması sonucunda elde edilen her bir kural (öğeler arasındaki birliktelik), destek (support) ve güven (confidence) kriterleri ile ifade edilir. Destek kriteri, öğeler arasındaki birlikteliğinin sıklığını, güven kriteri ise bu birlikteliklerin doğruluğunu ifade etmektedir.

A ve B, birbirinden farklı birer öge küme olsun. A öge kümesi için destek, A öge kümesini kapsayan kümelerin tüm öge kümelerine oranıdır ve (1) denklemi ile hesaplanır. A ve B öge kümeleri için destek ise tüm kümeler içerisinde birlikte bulunma olasılığıdır ve (2) denklemi ile elde edilir. B öge kümesinin hangi olasılıkla A öge kümeleri içerisinde bulunacağı güven değeri ile ifade edilir ve (3) ve (4) denklemlerinden biri ile hesaplanır.

$$\text{Destek (A)} = \frac{\text{A öge küme sayısı}}{\text{Toplam öge küme sayısı}} \quad (1)$$

$$\text{Destek (A,B)} = \frac{(A,B) \text{ öge küme sayısı}}{\text{Toplam öge küme sayısı}} \quad (2)$$

$$\text{Güven (A,B)} = \frac{(A,B) \text{ öge küme sayısı}}{\text{A öge küme sayısı}} \quad (3)$$

$$\text{Güven (A,B)} = \frac{\text{Destek(A,B)}}{\text{Destek (A)}} \quad (4)$$

Elde edilen kuralların kullanışlılığı, doğruluğu ve güvenilirliği, destek ve güven kriterleri ile belirlenir. İki öge kümenin birlikteliğinin önemli olması için hem destek

hem de güven kriterlerinin olabildiğince yüksek olması gerekmektedir.

Algoritmanın başlangıcında, kuralların geçerliliğini belirlemek amacıyla minimum destek ve güven kriterleri (eşik değerleri) belirlenir. Bu kriterler için optimum değerler seçilmesi önemlidir. Minimum destek kriteri büyük seçilirse apriori algoritmasının çalışma süresinin ve adımlarının azalmasının yanında elde edilen sonuç (kural) sayısı da azalacaktır. Böylece, elde edilen birliktelik kuralları kullanışlı olmayacaktır. Minimum güven kriterinin ise büyük seçilmesine özen gösterilmelidir. Çünkü güven kriteri, elde edilen kuralların doğruluğunu belirtmektedir [6].

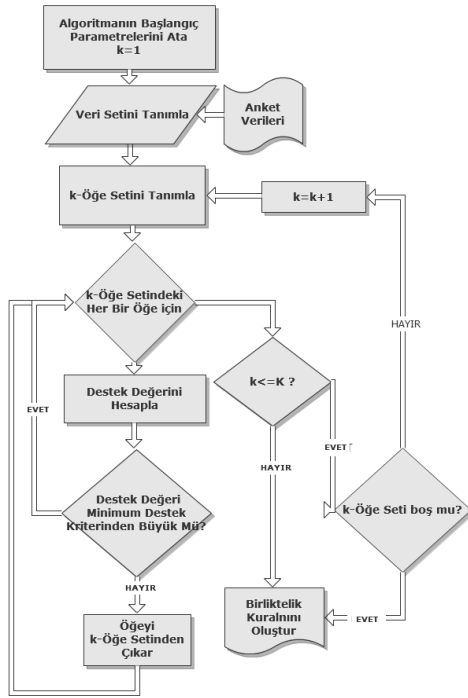
3. Uygulama

Bu çalışma kapsamında, Bilecik Şeyh Edebali Üniversitesi Bilgisayar Mühendisliği Bölümü 2., 3. ve 4. sınıf öğrencilerine Tablo 1' de yer alan 8 farklı teknik seçmeli ders (Teknik Seçmeli Ders 1, 2, ..., 8) seçim analizi ile ilgili anket yapılarak gerekli bilgiler alınmıştır. Bu bilgiler arasında ne tür ilişkilerin bulunduğu, bilgilerin nasıl dağılım gösterdiği, ders seçimleri ile seçilme nedenleri arasında ne gibi kuralların bulunduğu belirlenmeye çalışılmıştır.

Tablo 1. Teknik seçmeli dersler

Dersin kodu	Seçmeli ders paketindeki dersler	Teknik seçmeli dersin adı
S12_1_1	Bilgisayar Grafiğine Giriş	Teknik Seçmeli Ders 1
S12_1_2	Görsel Programlama	
S12_1_3	Biçimsel Diller ve Soyut Makineler	
S12_1_4	Bilg. Sistem ve Mod. ve Benzetim	
S12_2_1	Sistem Analizi ve Tasarımı	Teknik Seçmeli Ders 2
S12_2_2	Sayısal İşaret İşleme	
S12_2_3	Nöral Sistemlere Giriş	
S12_2_4	Yöneylem Araştırması	
S12_3_1	Yapay Zeka	Teknik Seçmeli Ders 3
S12_3_2	Bulanık Mantık	
S12_3_3	Görüntü İşleme	
S12_3_4	Makine Öğrenmesi	
S12_4_1	İleri Bilgisayar Mimarisi	Teknik Seçmeli Ders 4
S12_4_2	İleri Bilgisayar Grafikleri	
S12_4_3	İleri Bilgisayar Ağları	
S12_5_1	Gömülü Sistemler	Teknik Seçmeli Ders 5
S12_5_2	Mikroişlemci Uygulamaları	
S12_5_3	Robotik Sistemlere Giriş	
S12_6_1	Veri Madenciliği	Teknik Seçmeli Ders 6
S12_6_2	Sezgisel Optimizasyon	
S12_6_3	Derleyiciler	
S12_7_1	Kablosuz Ağ Teknolojileri	Teknik Seçmeli Ders 7
S12_7_2	Oyun Programlama	
S12_7_3	Çoklu Ortam Sistemleri	
S12_7_4	Ağ Güvenliği	
S12_8_1	DSP Uygulamaları	Teknik Seçmeli Ders 8
S12_8_2	FPGA Uygulamaları	
S12_8_3	Gömülü Sistem Tasarımı	

Matlab® programlama dili ile gerçekleştirilen apriori algoritmasının akış şeması Şekil 1' de verilmiştir.



Şekil 1. Apriori algoritmasının akış şeması

Anket yanıtları Excel tablosuna kaydedilmeden önce; eksik, gürültülü ve işe yaramayacak veriler göz ardı edilmiştir. Daha sonra yanıtlar "Ders kodu"- "Seçenek kodu" şeklinde kodlanarak Apriori algoritması için kullanıma uygun hale getirilmiş ve Excel tablosuna kaydedilmiştir. Örneğin, "Teknik Seçmeli 1" ders paketinden "Görsel Programlama" dersini "ders içeriği" sebebi ile seçen öğrencinin anket yanıtı " S12_1_2_b" şeklinde kodlanmıştır. Teknik seçmeli derslerin seçiminde kullanılan seçenekler ve seçenek kodları Tablo 2' de verilmiştir.

Tablo 2. Teknik seçmeli ders seçim nedenleri

Kodu	Seçenekler
a	Uygulaması mevcut
b	Ders içeriği
c	İlgi alanına yakınlık
d	Dersin zorluk derecesi
e	Dersi veren öğretim elemanı
f	Diğer(....)

Farklı minimum destek değerleri verilerek elde edilen Tablo 3' teki sonuçlara göre, en iyi sonuç % 20 destek değeri için elde edilmiştir. Tablo3' te görüldüğü üzere öğrenciler teknik seçmeli ders seçimlerinde, derslerin ilgi alanlarına yakın olup olmadıklarını dikkate almışlardır. %20 destek değerine göre elde edilen birliktelikler; "Sistem Analizi ve Tasarımı" ve "Yapay Zeka", "Sistem Analizi ve Tasarımı" ve "Robotik Sistemlere Giriş", "Yapay Zeka" ve "Robotik Sistemlere Giriş", "Yapay Zeka" ve "Gömülü Sistem Tasarımı", "Robotik Sistemlere Giriş" ve "Gömülü Sistem Tasarımı"dır.

Daha büyük destek değeri için birliktelik kuralı elde edilememiştir. %15 destek değeri için üçlü birliktelikler, %10 için dördü ve %5 için altılı birliktelikler elde edilmiştir. Elde edilen birlikteliklere bakıldığında seçimlerin belirli dersler üzerinde yoğunlaştığı görülmektedir. Yine Tablo 3' ten, teknik seçmeli 6 paketinden herhangi bir dersin farklı destek sayıları kullanılarak elde edilen kurallarda yer almadığı da görülmektedir. Bu durum, öğrencinin bu paketten ders seçiminin yeterli düzeyde olmadığı ya da o ders hakkında yeterli bilgisinin olmadığı söylenebilir.

Tablo 3. Farklı minimum destek değerleri ile elde edilen kurallar

Minimum Destek Kriteri	Kural Sayısı	Birliktelik Kuralları
%30	0	Birliktelik kuralı bulunmamaktadır.
%20	5	S12_2_1_c, S12_3_1_c S12_2_1_c, S12_5_3_c S12_3_1_c, S12_5_3_c S12_3_1_c, S12_8_3_c S12_5_3_c, S12_8_3_c
%15	4	S12_2_1_c, S12_3_1_c, S12_5_3_c S12_2_1_c, S12_5_3_c, S12_8_3_c S12_3_1_c, S12_5_3_c, S12_8_3_c S12_3_1_c, S12_5_3_c, S12_1_2_c
%10	2	S12_2_1_c, S12_3_1_c, S12_5_3_c, S12_8_3_c S12_2_1_c, S12_3_1_c, S12_5_3_c, S12_1_2_c
%5	3	S12_2_1_c, S12_3_1_c, S12_4_1_c, S12_5_2_c, S12_5_3_c, S12_7_4_c S12_2_1_c, S12_3_1_c, S12_5_3_c, S12_7_4_c, S12_4_3_c, S12_1_2_c S12_5_2_c, S12_5_3_c, S12_7_4_c, S12_4_3_c, S12_5_1_c, S12_7_1_c

4. Sonuç ve Öneriler

Bu çalışmada, birliktelik kuralı çıkarım algoritmalarından Apriori algoritmasının eğitimde kullanılmasına ilişkin bir uygulama gerçekleştirilmiştir. Elde edilen kurallar, öğrencilerin ders seçimlerinde hangi hususları göz önünde bulundurdıkları, hangi derslerin seçilme eğiliminin yüksek olduğu, hangi derslerin birlikte seçildiği gibi konularda bilgi içermektedir. Bu bilgi ile ilgili derslerin dönem başında açılması, işlenişinde neler yapılması gerektiği ve içeriğinin uygunluğu gibi düzenlemelere gidilmesi için bir ön hazırlık niteliğindedir.

Burada dikkate alınan algoritma; öğrencileri yakından ilgilendiren erasmus / farabi, yaz okulu, staj gibi eğitim alanındaki diğer konularda da birlikteliklerin bulunmasında kullanılabilir.

5. Kaynaklar

- [1] Özmen, Ş., İş Hayatı Veri Madenciliği ile İstatistik Uygulamalarını Yeniden Keşfediyor, <http://idari.cu.edu.tr/sempozyum/bil38.htm> [11.12.2012]
- [2] Küçüksille, E., Veri Madenciliği Süreci Kullanılarak Portföy Performansının Değerlendirilmesi ve IMKB Hisse Senetleri Piyasasında Bir Uygulama, **Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü**, Doktora Tezi, Isparta (2009)
- [3] Ay, D. ve Çil, İ., Migros Türk A.Ş.De Birliktelik Kurallarının Yerleşim Düzeni Planlamada Kullanılması, **Endüstri Mühendisliği Dergisi**, cilt: 21, sayı: 2, 14-29 (2008).

[4] Chen, Y.L., Chen, J.M. ve Tung, C.W., A Data Mining Approach For Retail Knowledge Discovery With Consideration of the Effect of Shelf-Space Adjacency on Sales, **Decisions Support Systems**, cilt: 42, sayı: 3, 1503-1520 (2006).

[5] Agrawal, R. ve Srikant, R., Fast Algorithms for Mining Association Rules in Large Databases, **Proceedings of the 20th International Conference on Very Large**

Databases (VLDB), Santiago, 487-499 (1994).

[6] Özseven, T. ve Düğenci M., LOG Analiz: Erişim Kayıt Dosyaları Analiz Yazılımı ve GOP Üniversitesi Uygulaması, **Bilişim Teknolojileri Dergisi**, cilt: 4, sayı: 2, 55- 66 (2011).

[7] Han, J. ve Kamber, M., Data Mining: Concepts and Techniques, **Morgan Kaufmann Publishers**, San Francisco, 2001.