

VERİ MADENCİLİĞİNDE APRIORİ ALGORİTMASI VE APRIORİ ALGORİTMASI KULLANARAK SÖZ KONUSU ÜRÜNLERİN BİRLİKTE SATILMASI DURUMLARININ İNCELENMESİ

Gizem Arslan, Fırat Üniversitesi, Teknoloji Fakültesi, Yazılım Mühendisliği
Bölümü, 23000, Elazığ, TÜRKİYE

Özet: Apriori algoritması, veri madenciliğinde sık geçen öğelerin keşfedilmesi için kullanılan en çok bilinen birliktelik-ilişki kuralı algoritmasıdır. Günümüzde teknoloji sayesinde çok büyük miktarda veri elde edilip saklanabilmektedir. Veri madenciliği, büyük miktardaki mevcut veri içinden anlamlı, potansiyel olarak kullanışlı, gelecekle ilgili tahmin yapılmasını sağlayan bağıntı ve kuralların bilgisayar programları kullanarak bulunmasıdır. Bu çalışmada elde edilen veri seti üzerindeki ürünlerin hangi sıklıkla alındığı hangi ürünün hangi ürünle birlikte alındığı gibi sorulara yanıtlar bulacağız. Bu çalışmada başarılı sonuçlar elde etmek için kullanılacakım verileri elde etmek için aktif bir mobilya mağazasının belirli süre zarfında sattığı ürünleri ve ürünleri satış sıklığını inceledim. Veri setini kendim oluşturdum. Çalışmalar bittikten sonra elde edilen veriler weka yazılıma gönderilip ve birliktelik kurallarının çıkarılması sağlanacaktır.

Anahtar kelimeler: Veri madenciliği, Apriori algoritması, Birliktelik kuralı

1. GİRİŞ

Veriler belli amaçlar doğrultusunda kullanıldığında bir anlam ifade ederler. Veri madenciliğinde kullanılan ilk metotlardan biri olan birliktelik kuralı geçmiş verilerin analiz edilerek bu veriler içindeki birliktelik davranışlarının tespiti ile geleceğe yönelik çalışmalar yapılmasını destekleyen bir yaklaşımdır[1]. 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 oldukça faydalıdır [2,3].

Bu çalışmada anlaşılmalı bir mağazanın gerçek veri seti üzerinden tanımlayıcı tekniklerden biri olan birliktelik kuralının çıkarılması amaçlanmıştır.

2. BİRLİKTELİK KURALI

Olayların birlikte gerçekleşme durumlarını çözümleyen veri madenciliği yöntemlerine birliktelik kuralları (association rules) adı verilmektedir. Bu yöntemler birlikte olma kurallarını belirli olasılıklarla ortaya koyar.

Birliktelik çözümlerinin en yaygın uygulaması perakende satışlarda müşterilerin satın alma eğilimlerini belirlemek amacıyla yapılmaktadır. Müsterilerin bir anda satın aldığı tüm ürünleri ele alarak satın alma eğilimini ortaya koyan uygulamalara “Pazar sepet çözümlemesi” adı verilmektedir[4].

2.1 Destek ve Güven Ölçütleri

Pazar sepet çözümlemelerinde satılan ürünler arasındaki ilişkileri ortaya koymak için “**destek**” ve “**güven**” gibi iki ölçütten yararlanılır. Bu ölçütlerin hesaplanmasında “**destek sayısı**” adı verilen bir değer kullanılır. “**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 grubunda alma durumu, yani birliktelik kuralı $A \rightarrow B$ biçiminde gösterilir. Bu durumda kural destek ölçütü şu şekilde ifade edilebilir:

$$\text{destek}(A \rightarrow B) = \text{sayı}(A, B) / N$$

Burada $\text{sayı}(A, B)$ destek sayısı A ve B ürün gruplarını birlikte içeren alışveriş sayısını içermektedir. A ve B ürün gruplarının birlikte satın alınması olasılığını ifade eden *kural güven ölçütü* şu şekilde hesaplanır[Tan, 2006].

$$\text{güven}(A \rightarrow B) = \text{sayı}(A, B) / \text{sayı}(A)$$

Birliktelik kuralları belirlenirken yukarıda söz edilen destek ve güven ölçütleri yanı sıra, bu değerleri karşılaştırmak üzere eşik değere gereksinim vardır. Hesaplanan destek veya güven ölçütlerinin *destek(eşik)* ve *güven(eşik)* değerlerinden büyük olası beklenir. Hesaplanan destek veya güven ölçütleri ne kadar büyük ise birliktelik kurallarının da o derece güçlü olduğuna karar verilir[4].

3. APRIORİ ALGORİTMASI

Birliktelik kuralının üretilmesi için bir çok yöntem kullanılmaktadır. Bunlardan en yaygın kullanılanı “Apriori algoritması” adını taşımaktadır. Bu algoritma aşağıda belirtilen aşamalara sahiptir.

- Birliktelik çözümlemesinin yapılabilmesi için öncelikle destek ve güven ölçütlerini karşılaştırmak üzere eşik değerler belirlenir. Uygulamadan elde edilen sonuçların bu eşik değerlere eşit yada büyük olması beklenir.
- Veritabanı taranarak çözümlemeye dahil edilecek her bir ürün için tekrar sayıları, yani destek sayıları hesaplanır. Bu destek sayıları eşik destek sayısı ile karşılaştırılır. Eşik destek sayısından küçük değerlere sahip satırlar çözümlemiden çıkarılır ve koşula uyum kayıtlar göz önüne alınır.

- c) Yukarıda adımda seçilen ürünler bu kez ikişerli gruplandırarak, bu grupların tekrar sayıları, yani destek sayıları elde edilir. Bu sayılar eşik destek sayıları ile karşılaştırılır. Eşik değerden küçük değerlere sahip satırlar çözümlemenden çıkartılır.
- d) Bu kez üçerli dörderli vb gruplandırmalar yapılarak bu grupların destek sayıları elde edilir ve eşik değer ile karşılaştırılır, eşik değerlere uygun olduğu sürece işlemlere devam edilir.
- e) Ürün grubu belirlendikten sonra kural destek ölçütüne bakılarak birliktelik kuralları türetilir ve bu kuralların her birisiyle ilgili olarak güven ölçütleri hesaplanır[4].

4. VERİLERİN TOPLANMASI

Bu çalışmada Apriori algoritması kullanılarak müşteriler bir ürün aldıklarında satış elemanlarının ‘Bu ürünün yanında hangi ürünü satabiliriz?’ sorusuna yanıt verecek kurallar çıkartılmıştır. Bir kaç inceleme yaptıktan sonra kullanılan veriler oluşturulmuştur.

5. SONUÇ VE YORUM

Kurallar minimum yüzde 90 güvenirlilikte ve 5 adettir. Aşağıdaki çıkarılan birliktelik kuralları gösterilmiştir.

1. kitaplik=aldi tv_unitesi=aldi yemek_masasi=aldi 16 ==> koltuk_takimi=aldi 16
<conf:(1)> lift:(1.8) lev:(0.06) [7] conv:(7.13)
2. koltuk_takimi=almadi tv_unitesi=aldi yemek_masasi=aldi 14 ==> kitaplik=almadi 14
<conf:(1)> lift:(2.42) lev:(0.06) [8] conv:(8.2)
3. koltuk_takimi=almadi kitaplik=almadi yemek_masasi=aldi 15 ==> tv_unitesi=aldi 14
<conf:(0.93)> lift:(1.9) lev:(0.05) [6] conv:(3.81)
4. koltuk_takimi=almadi kitaplik=almadi tv_unitesi=aldi 15 ==> yemek_masasi=aldi 14
<conf:(0.93)> lift:(1.66) lev:(0.04) [5] conv:(3.28)
5. koltuk_takimi=aldi yemek_odasi=almadi komidin=almadi 14 ==> yemek_masasi=aldi 13
<conf:(0.93)> lift:(1.65) lev:(0.04) [5] conv:(3.06)

Kural 3’e göre yüzde 93 doğruluk ile koltuk takımı ve kitaplık almayan müşteriler yemek masası aldığı anda tv üniteside alıyor. Bu durumda satış elemanlarının söz konusu ürünleri satarken yapacağı işlem koltuk takımı ve kitaplık almayan fakat yemek masası alan müşterilere tv ünitesi almayı önermektir.

6. KAYNAKLAR

- [1] <https://www.veribilimi.co/associationrulesanalysis/> , 19.05.2019
- [2] Ay, D. ve Çil, İ., (2008). Migros Türk A.Ş.De Birliktelik kurallarının yerleşim düzeni planlamada kullanılması, Endüstri Mühendisliği Dergisi, 21(2), 14-29.
- [3] Güngör, E., Yalçın, N. Ve Yurtay, N., (2013). Apriori Algoritması ile Teknik Seçmeli Ders Seçim Analizi. Ulusal Uzaktan Eğitim ve Teknolojileri Sempozyumu, 1-3.
- [4] Dr.Yalçın Özkan,Veri Madenciliği Yöntemleri, İstanbul, Papatya yayınları, 2016, 217.