



Case Study

A SOFTWARE DESIGN AND APPLICATION FOR APRIORI ALGORITHM: A CASE STUDY ON EVALUATION OF ISKUR DATA

APRİORİ ALGORİTMASININ KULLANILMASINA YÖNELİK BİR YAZILIM TASARIMI VE UYGULAMASI: İŞKUR VERİLERİNİN DEĞERLENDİRİLMESİ ÜZERİNE BİR ÖRNEK ÇALIŞMA

Doruk AYBERKİN^{1,*} Üstün ÖZEN²

¹ Öğr. Gör., Bayburt TBMYO, Bilgisayar Teknolojileri ve Programlama Bölümü, Türkiye, doruk@bayburt.edu.tr

² Prof. Dr., Atatürk Üniversitesi İİBF Yönetim Bilişim Sistemleri Bölümü, Türkiye, uozen@atauni.edu.tr

Article Info:

Received : December 6, 2019

Revised : December 23, 2019

Accepted : December 26, 2019

Keywords:

Data Mining

Association Rule Mining

Apriori Algorithm

Turkish Employment Agency

Anahtar Kelimeler:

Veri Madenciliği

Birliktelik Kuralı Madenciliği

Apriori Algoritması

Türkiye İş Kurumu

ABSTRACT

Data mining is used to reveal relationships between variables stored in large databases and to reach meaningful information from data. Databases with large data available in government agencies can be evaluated with appropriate software to help predict the future and develop policies. In this study, a software that analyzes data using apriori algorithm is designed. In the implementation phase of the Apriori algorithm, the study was extended to take into account the measure of lift in addition to the trust and support criteria. After analyzing the data of Turkish Employment Agency imported into the software, a set of rules specific to job applications were created and it's finally interpreted.

ÖZET

Veri madenciliği, büyük veri tabanlarında depolanan değişkenler arasındaki ilişkileri ortaya çıkarmak ve verilerden anlamlı bilgilere ulaşmak için kullanılır. Devlet kurumlarında bulunan büyük boyutlu veriye sahip veri tabanları uygun yazılımlarla değerlendirilerek gelecek ile ilgili tahminlerin yapılabilmesine, politikalar geliştirilmesine yardımcı olabilir. Bu çalışmada Apriori algoritmasını kullanarak verileri analiz eden bir yazılımın tasarımı yapılmıştır. Apriori algoritmasının uygulama aşamasında güven ve destek kriterlerine ek olarak ilginçlik ölçüsü de hesaba katılarak çalışma genişletilmiştir. Yazılım içerisine aktarılan Türkiye İş Kurumu'na ait verilerin analizi yapılmış, iş başvurularına özgü kurallar kümesi oluşturularak yorumlanmıştır.

© 2019 JOBDA All rights reserved

1 | GİRİŞ

İnternet ve bilişim teknolojilerinin gelişimi ile birlikte sayısal verilerde büyük bir artış yaşanmıştır. Toplanan sayısal veriler detaylı ve amaca uygun biçimde sayısal veri tabanlarında tutulmakta, istenildiğinde süratle kullanılabilir durumda hazır bulunmaktadır. Veri işlenmemiş olarak tek başına bir

anlam ifade etmez, ancak veriler üzerinde analiz yaparak bir sorunun karşılığı olan bilgiye ulaşılabilir.

Ortaya çıkan büyük ham veriden istenilen çözüme uygun bilgi üretilmesi süreci Veri Tabanından Bilgi Keşfi (VTBK-Knowledge Discovery in Databases) olarak adlandırılmaktadır. VTBK yaklaşımı; veri temizleme, veri birleştirme, veri dönüştürme, örüntü

* Corresponding Author,

E-mail: doruk@bayburt.edu.tr (Doruk Ayberkin)

tanımlama ve yorumlama ile bilgi sunumunu içermektedir. Verilerin anlamlı bilgi haline dönüştürülmesi istendiğinde çözüm olarak Veri Madenciliği (Data Mining) kavramı devreye girmektedir.

Veri madenciliği uygulamaları, veriden anlamlı bilgiler çıkarmak için günümüzde birçok alanda kullanılmaktadır. Birliktelik kuralları ise veri madenciliğinde kullanılan bir teknik olup, veri tabanlarındaki bilgiler arasında anlamlı ilişkileri ortaya çıkararak her verinin birbiri ile ilişkisini açıklamaktadır. Daha çok market sepeti analizi olarak bilinse de ilişki kurgusu olan verilerin tamamında kullanılabilen bir yöntemlerdir. Bu analizde, Apriori algoritması gibi farklı algoritmalar da kullanılmaktadır.

Bu çalışmada ilişkili olacağı düşünülen Türkiye İş Kurumu (İŞKUR) iş başvurusu verilerinin içerisinde, birliktelik kuralı bulmak amacıyla, Apriori algoritması kullanılarak inceleme yapılmıştır. Bulunan birliktelik kuralları yorumlanmış ve çıkan sonuçlar tartışılmıştır. Sonuç olarak, başvuru verileri üzerinden ulaşılan anlamlı bilgiler sayesinde İŞKUR'un ve bağlantılı diğer kurumların, fayda sağlaması amaçlanmıştır.

2 | KAVRAMSAL ÇERÇEVE

2.1 | VERİ MADENCİLİĞİ VE BİRLİKTELİK KURALLARI

2.1.1 | Veri Madenciliğinin Tanımı ve Kullanılan Modeller

Veri madenciliği modelleri temelde iki ana başlık altında incelenmektedir. İlk başlık sonuçları belirli olan verileri kullanarak gelecekteki davranışları belirleyebileceğimiz tahmin edici, diğer başlık ise verilerdeki örüntülerin bulunmasını sağlayarak karar vermeye yardımcı olan tanımlayıcıdır (Fayyad vd.1996, s. 85). Tahmin edici başlığı altında sınıflama ve regresyon modelleri, tanımlayıcı başlığı altında ise kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri bulunmaktadır (Akpınar, 2000, s. 2).

2.1.1.1 | Birliktelik Kuralları

Birliktelik kuralları tanımlayıcı veri madenciliğinde yaygın olarak kullanılan ve denetimsiz öğrenen veri madenciliği modellerinden biridir. Bu kurallar, veriler arasındaki ilişkileri tanımlayarak birlikte olma durumlarını inceler. Bu sayede veri kümesi içerisinde bulunan eş zamanlı meydana gelme durumları ve oluşumlar incelenebilmektedir. Bir başka deyişle, birliktelik kuralları veri depolarındaki bağlantısız veriler arasında birliktelik ve ilişkileri arayarak sıkça kullanılan nesnelere bağlantılarını ortaya çıkartmak için kullanılır (Kumbhare & Chobe, 2014, s. 927).

Birliktelik kuralları ilk olarak 1993 yılında Agrawal ve arkadaşları (Agrawal vd.1993, s. 207) tarafından bir

süpermarkette müşterilerin aldığı ürünler arasındaki birliktelikleri ortaya koymak amacıyla kullanılmıştır. Bu nedenle literatürde genellikle market sepeti analizi adıyla kullanılmaktadır. Birliktelik kurallarının matematiksel modeli aşağıdaki şekilde tanımlanmıştır.

$I = \{i_1, i_2, \dots, i_m\}$, farklı değişkenlere sahip m öğeden oluşan bir veri kümesidir. Veri kümesinde t , küme içindeki işlemler ve k , kurallar olmak üzere birliktelik olması durumunda $t[k] = 1$ aksi halde $t[k] = 0$ olur. Bir birliktelik kuralı şeklinde bir ifade edilir. Burada X ve Y , tam veri kümesi olan I 'nin alt kümeleridir. kuralı, X olduğunda Y 'nin de olduğunu, yani X ve Y 'nin birlikte oluştuğunu, kısaca birlikteliğini ifade etmektedir (Agrawal vd.1993, s. 208). Daha anlaşılır şekilde kuralın genel tanımı, X ' teki öğeleri içeren veri tabanındaki işlemlerin Y 'deki öğeleri de içermeye eğilimi göstermesidir (Agrawal & Shafer, 1996, s. 963).

Birliktelik kuralları analizinde ortaya birçok ilişki kuralı çıkmakta, bunlar arasından seçim yapılması aşamasında, destek ve güven ölçütleri devreye girmektedir. Destek ölçütü istatistiksel bir boyut olarak görev görür; yeterince büyük değerse kuralın daha az tercih edildiği ve dikkate alınmayabileceğini gösterir. Bu durum belirlenen asgari eşik değeri göz önüne alınarak değerlendirilir. Güven ölçütü ise tercih edilen durumların hangi olasılıkla gerçekleştiği hakkında bilgi vermektedir. Destek ve güven ölçütleri sırasıyla eşitlik (1) ve eşitlik (2)' de ifade edilmiştir (Agrawal & Srikant, 1994, s. 487).

$$\text{Destek} = \frac{\text{X ve Y'nin beraber bulunduğu satır sayısı}}{\text{Toplam satır sayısı}} \quad (1)$$

$$\text{Güven} = \frac{\text{X ve Y'nin beraber bulunduğu satır sayısı}}{\text{X'in bulunduğu satır sayısı}} \quad (2)$$

Destek ve güven ölçütleri için belirlenecek olan minimum değerler, çalışmayı yapan araştırmacı tarafından belirlenir. Ancak unutulmamalıdır ki birliktelik kuralları, belirlenen minimum değerler ile anlam kazanmaktadır. Belirlenen değerlerin çok düşük olması ilişki sayısının fazla, çok yüksek olması ise ilişki sayısının az çıkmasına neden olacaktır. Minimum destek değeri yaygın olarak bulunan eşleşmeleri belirlerken, minimum güven değeri bu eşleşmeler içerisinde belirlenen minimum değere göre elemeler yaparak güçlü birlikteliklerin ortaya çıkarılmasını sağlamaktadır.

Birliktelik kurallarında seçimi etkileyen diğer bir faktör ise ilginçlik ölçüsü olarak adlandırılan Lift değeridir. Lift değeri genel olarak nitelikler arasındaki ilişkinin yoğunluğunu ifade eder; daha büyük Lift değeri, nitelikler arasında daha yoğun bağlantı anlamına gelir (Linoff & Berry, 2011, s. 81). Lift değeri, bulunan kurallar içerisinde bir kuralın diğerlerinden ne derece önemli olduğunu belirlemek

için kullanılır. Lift ölçütü eşitlik 3 'te ifade edilmiştir (Hastie vd.2009, s. 491).

$$\text{Lift}(X \rightarrow Y) = \frac{\text{Destek}(X \text{ ve } Y)}{\text{Destek}(X) \cdot \text{Destek}(Y)} \quad (3)$$

Birliktelik kuralına ilişkin olarak geliştirilen bazı algoritmalar şunlardır; AIS (Agrawal vd.1993, s. 207-216), Apriori (Agrawal & Srikant, 1994, s. 487-499), SETM (Houtsma & Swami, 1995, s. 245-262), RARM (Das vd.2001, s. 474-481), CHARM (Zaki & Hsiao, CHARM: An Efficient Algorithm for Closed Itemset Mining). Bu algoritmalar içerisinde ilk olanı AIS, en bilineni ise Apriori algoritmasıdır.

2.1.1.2 | Apriori Algoritması

Apriori 1994 yılında Agrawal ve Srikant tarafından veriler arasındaki mantıksal birliktelik kurallarını kullanarak veri madenciliği yapmak için önerilen bir algoritmadır (Agrawal & Srikant, 1994, s. 487-499). Algoritmanın adı, sık kullanılan öge kümesi özellikleri hakkında önceki (prior) bilgileri kullanmasına dayanmaktadır.

Apriori, birliktelikleri keşfetmek için hazırlanmış yinelemeli bir yaklaşım kullanmaktadır. İlk olarak, her bir ögenin toplam sayısını bulmak için veri tabanını tarar ve minimum desteği sağlayan öğeleri bir araya getirerek, bir öge kümesi bulur. Elde edilen küme daha sonra alt kümeleri bulmak için kullanılır ve tüm birliktelikler bulunana kadar bu işlem devam eder. Her birliktelik alt kümesinin bulunması, veri tabanının bir tam taramasını gerektirir. Apriori algoritması bu taramalar sırasında verimliliği arttırmak ve arama alanını azaltmak için Apriori özelliği adı verilen bir hesaplama kullanır (Han, Kamber, & Pei, 2011). Bu özelliğe göre, bir öge kümesi yaygın bir öge kümesi ise bu öge kümesinin tüm alt kümeleri de yaygın öge kümesi olmalıdır (Salvatore vd.2004, s. 542).

Apriori algoritması iki aşamadan oluşan bir süreçtir. Birinci aşamada birleştirme ikinci aşamada ise budama işlemi gerçekleştirilir. Sık geçen k ögeli birliktelik kümeleri L_k , k aday altküme sayısı ve k ögeli aday birliktelik kümeleri C_k olarak adlandırılır. Birleştirme adımında L_k 'yı bulmak için k ögeli her bir aday küme (C_k), L_{k-1} 'in kendi arasında birleştirilmesiyle oluşturulur. C_k 'nın elemanları bütün birliktelikleri içerir ve L_k 'nın bir üst kümesidir. C_k 'daki aday sayısını belirlemek için bir veri tabanı taraması yapılır ve L_k 'lar hesaplanır. C_k sayısı çok büyük olduğu durumlarda ağır hesaplamalar oluşacağından C_k sayısını azaltmak için budama aşaması gerçekleştirilir. Budama aşamasında, yaygın olmayan herhangi bir (k-1) aday kümesi bir yaygın k-1 aday kümesinin alt kümesi olamaz ve eğer bir aday k-1 alt kümesinin herhangi bir (k-1) alt kümesi L_{k-1} 'de yoksa aday yaygın aday kümesi olamayacağından ve C_k 'dan kaldırılır. Bu sayede yaygın olmayan kurallar

budanarak yaygın olan kuralların kalması sağlanır (Han vd.2011).

Apriori algoritmasının klasik sözde kodu, Şekil 1. (Han vd.2011, s. 253)' de görülmektedir. Şekilde yer alan apriori-gen fonksiyonu, (k-1) adet ögeye sahip L_{k-1} sık geçen aday kümesini kullanarak k adet ögeye sahip aday alt kümelerin oluşturulmasını sağlar. Bu fonksiyon ile ilk olarak, L_{k-1} sık geçen öge kümesine kendisi ile birleştirme işlemi uygulanır. Birleştirme işleminde L_{k-1} sık geçen öge kümesinin her satırında yer alan son öge haricinde diğer öğelerin çapraz olarak benzerliği aranır ve son öge haricinde diğer öğelerle yakalanan benzerliklerden yeni aday öge kümeleri oluşturulur. Oluşan kümeler budama adımı ile budanarak fonksiyondan ana işleme dönülür (Agrawal & Srikant, 1994, s. 491).

Şekil 1. Apriori Algoritmasının Sözde Kodları

```

1)  $L_1 = \{ \text{sık geçen 1-öge kümesi} \};$ 
2) for ( $k=2; L_{k-1} \neq \emptyset; k++$ ) do begin
3)    $C_k = \text{apriori-gen}(L_{k-1});$  // Ycni adaylar
4)   for all transactions-hareketler  $t \in D$  do begin
5)      $C_t = \text{subset}(C_k, t);$  // Adaylar t içindedir
6)     for all candidates - adaylar  $c \in C_t$  do
7)        $c.\text{count}++;$ 
8)   end
9)    $L_k = \{ c \in C_k \mid c.\text{count} \geq \text{minsup} \}$ 
10) end
11) Answer =  $\cup_k L_k;$ 

```

3 | UYGULAMA

3.1 | APRIORI ALGORİTMASININ UYGULANMASINA YÖNELİK YAZILIM UYGULAMASI

3.1.1 | Yazılım Ortamı ve Verilerin Hazırlanması

Uygulama yazılımı Microsoft.NET platformunda ve C# programlama dili kullanılarak hazırlanmıştır. Uygulama içerisinde oluşturulan kütüphaneler ile Apriori algoritması çalıştırılmaktadır.

Uygulamada Türkiye İş Kurumu (İŞKUR)'dan sağlanan 2017 yılına ait veriler kullanılmıştır. Veriler Microsoft Excel hesap tablosu biçiminde bulunmaktadır. Veriler, algoritmanın çalıştırılması için uygulama içerisinde hazırlanarak uygun hale getirilmektedir. Verilerin hazırlanması süreci VTBK'nın veri seçimi, veri ön işleme ve veri indirgeme aşamalarından oluşmaktadır. (Fayyad vd.1996, s. 84)

Veri seçimi sürecinde İŞKUR'dan alınan on aya ait başvuru verileri tek bir tabloda toplanmıştır. Daha sonra toplanan veriler içerisinde yer alan ancak kullanımı bir anlam taşımayan diğer bir ifadeyle gürültülü veri olarak adlandırılan (Han vd.2011) alanlar tespit edilerek, birliktelik kurallarının ortaya çıkarılmasına hizmet edecek biçimde yeniden düzenlenmiştir. Ön işleme ve indirgeme sürecinden geçirilen veriler üzerinde geliştirilen uygulama vasıtasıyla Apriori algoritması uygulanmıştır.

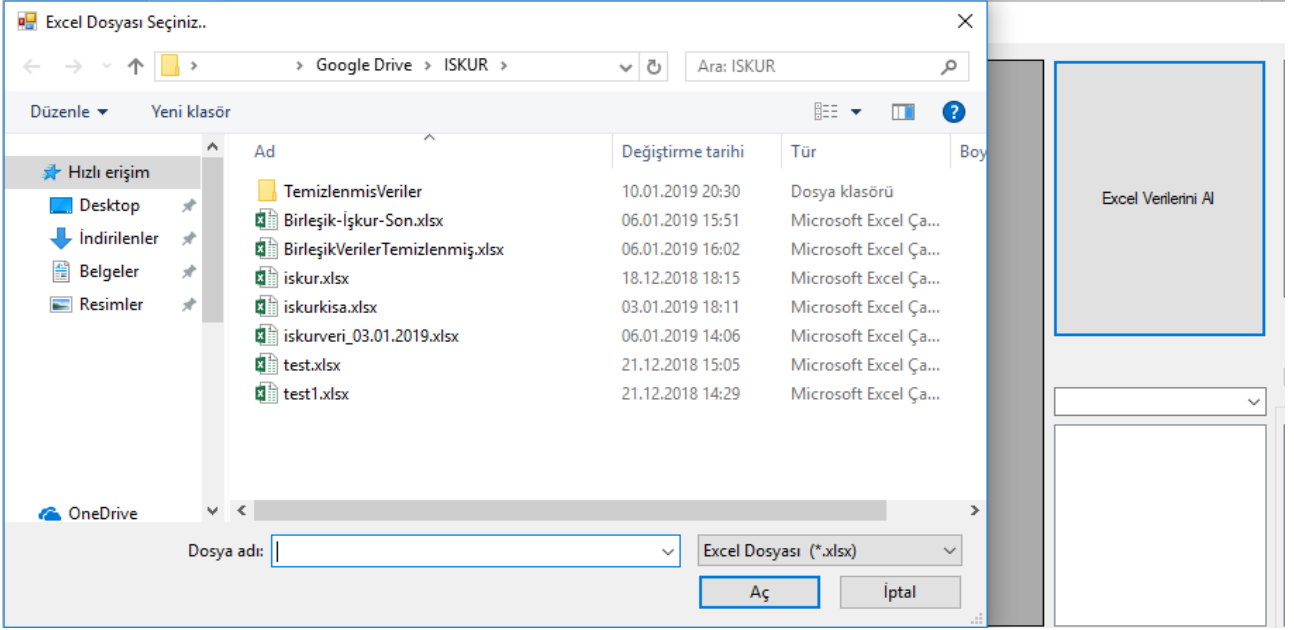
Geliştirilen uygulama yazılımı çalışma aşamaları ve yapılan işlemlerin açıklaması izleyen başlıklarda verilmiştir.

3.1.2 | Verilerin Yazılım İçerisine Alınması

İŞKUR'dan Excel uzantısı (.xlsx) biçiminde elde edilen aylık veriler ilk olarak birleştirme işlemine tabi tutulmuştur. Birleştirme işlemi, veri bütünlüğünü sağlamak ve sonraki aşamalarda yapılacak hesaplamaların güvenilirliğini arttırmak açısından

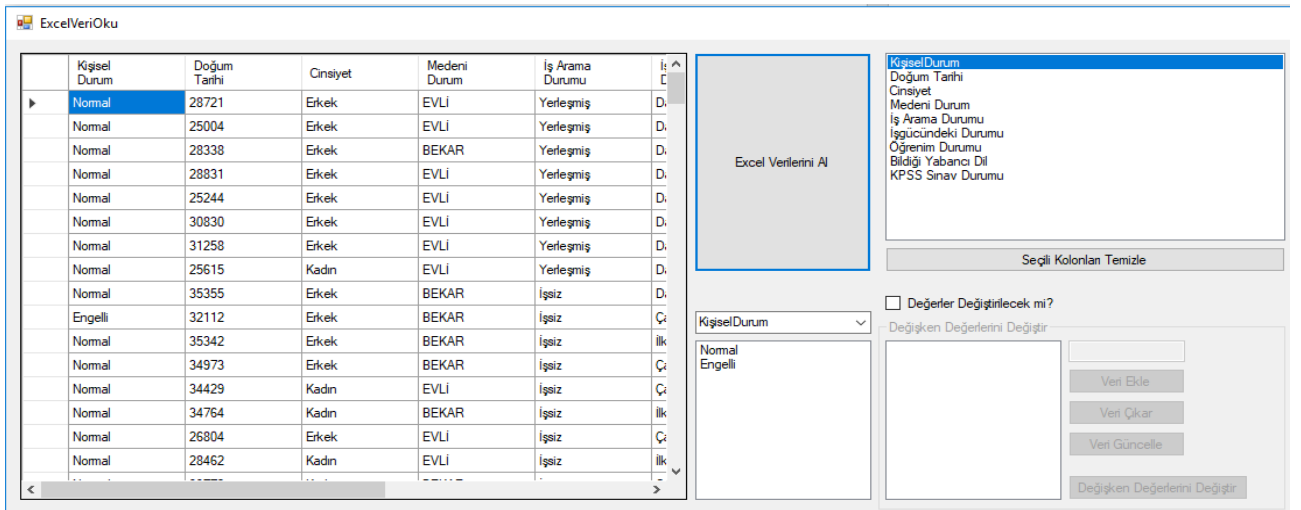
önem taşımaktadır. Birleştirme aşamasından sonra verilerin daha rahat işlenebilmesi için program içerisine alınma süreci başlamaktadır. Bu aşamada verilerin hazırlanmış olduğu xlsx dosya formatı XML biçimli bir dosya ürettiğinden, yazılım bu yapıyı destekler bir yapıda hazırlanmıştır. Veriler program içerisinden dosya yolu gösterilerek veriler yazılım içerisinde tablo formatına dönüştürülür. Bu işlem Şekil 2' de gösterilmiştir

Şekil 2. Verilerin Yazılım İçerisine Alınması



Veriler, yazılım içerisine alındığında bir tablo elemanı içerisine doldurulmakta aynı zamanda veri dosyası içerisinde ilk satırda bulunan alan isimleri de bir liste elemanı içerisinde saklanmaktadır. Bu sayede hem bir sonraki aşamada yapılacak olan sonuç odaklı olmayan alanların çıkarılması hem de veri dosyasında bir hata oluşması durumunda tespiti sağlanmıştır. Verilerin yazılım içerisine alınması sonrası oluşan yazılım görüntüsü Şekil 3' te gösterilmiştir.

Şekil 3. Veriler Yazılım İçerisine Alındıktan Sonra Oluşan Görüntü



3.1.3 | Sonuç odaklı olmayan alanların veri tabanından çıkarılması

Bu süreçte, önemsiz alanlar çıkartılarak veri seti filtrelenir. Bu aşamada amaç algoritmanın uygulanması sırasında anlamsız birlikteliklerin çıkmasını engelleyerek birlikteliklerin bulunma süresini kısaltmaktır. Örnek vermek gerekirse veri setinde bulunan askerlik durumu bilgisi, erkek adaylar için anlam taşırken kadın adaylar için anlamsız bir alandır. Veri seti erkek ve kadın adayları bir bütün

olarak barındırdığından, askerlik durumu anlamlı birliktelik kuralına uymadığından veri setinden çıkartılmıştır.

Veri setinden alan çıkarma işlemi, listelenmiş başlıklar içerisinde seçim yapılarak gerçekleştirilir. Seçim işlemi tek bir eleman veya birden fazla eleman için aynı anda gerçekleştirilir. Çıkartma işlemi Şekil 4' te gösterilmiştir.

Şekil 4. Veri Setinden Alan Eksiltme İşlemi

The screenshot shows the ExcelVeriOku application interface. On the left, there is a table with columns: Kişisel Durum, Doğum Tarihi, Cinsiyet, Medeni Durum, İş Arama Durumu, and a dropdown menu. The table contains 20 rows of data. On the right, there is a panel titled 'Excel Verilerini Al' with a list of fields: Kişisel Durum, Doğum Tarihi, Cinsiyet, Medeni Durum, İş Arama Durumu, İşgücündeki Durumu, Öğrenim Durumu, Bildiği Yabancı Dil, and KPSS Sınav Durumu. Below this list is a button 'Seçili Kolonları Temizle'. A dialog box titled 'Uyarı' is open, asking 'Veri setinden seçili alanları çıkarmak istediğinize emin misiniz?' with 'Tamam' and 'İptal' buttons. Below the dialog box are buttons for 'Veri Güncelle' and 'Değişken Değerlerini Değiştir'.

Kişisel Durum	Doğum Tarihi	Cinsiyet	Medeni Durum	İş Arama Durumu	
Normal	28721	Erkek	EVLİ	Yerleşmiş	D.
Normal	25004	Erkek	EVLİ	Yerleşmiş	D.
Normal	28338	Erkek	BEKAR	Yerleşmiş	D.
Normal	28831	Erkek	EVLİ	Yerleşmiş	D.
Normal	25244	Erkek	EVLİ	Yerleşmiş	D.
Normal	30830	Erkek	EVLİ	Yerleşmiş	D.
Normal	31258	Erkek	EVLİ	Yerleşmiş	D.
Normal	25615	Kadın	EVLİ	Yerleşmiş	D.
Normal	35355	Erkek	BEKAR	İşsiz	D.
Engelli	32112	Erkek	BEKAR	İşsiz	Ç.
Normal	35342	Erkek	BEKAR	İşsiz	İlk
Normal	34973	Erkek	BEKAR	İşsiz	Ç.
Normal	34429	Kadın	EVLİ	İşsiz	Ç.
Normal	34764	Kadın	BEKAR	İşsiz	İlk
Normal	26804	Erkek	EVLİ	İşsiz	Ç.
Normal	28462	Kadın	EVLİ	İşsiz	İlk

3.1.4 | Nitel Verilerin Nicel Değerlere Dönüştürülmesi

Apriori algoritması, niteliklerin sahip olduğu var/yok diğer bir ifadeyle bir ve sıfır değerleri ile işlem yapmaktadır. Ancak niteliklerin aldığı değerler incelenen problemlerin birçoğunda ikiden fazla olabilmektedir. Bu gibi durumlarda, niteliklerin sayısal veya kategorik değerleri ile kural üretmek için Srikant ve Agrawal tarafından 1996 yılında yapılan çalışmada etkili bir metot önerilmiştir. Önerilen metot, niteleyici aralık değişkenlerinin değer sırasının korunarak ardışık tam sayılara eşlenmesine

dayanmaktadır (Srikant & Agrawal, 1996, s. 1-12). Bu nedenle yazılım içerisinde Apriori algoritmasının çalıştırılması için nitel değerler nicel değerlere dönüştürülmüştür.

Yazılımda alan başlıklarının doldurulduğu bir liste elemanına bağlı olarak alt nitelikler listelenmektedir. Değişiklikler yazılım içerisinde bulunan "Değişken değerlerini değiştir" alanından niteliğe karşılık sayısal bir değer verilerek gerçekleştirilir. Değer değiştirme işlemi ve örnek ön izleme Şekil 5' te gösterilmiştir.

Şekil 5. Nitel Değişkenlerin Sayısal Biçime Dönüştürülmesi

The screenshot shows the ExcelVeriOku interface. On the left, a data table is displayed with columns: Cinsiyet, Medeni Durum, İş Arama Durumu, İlgücündeki Durumu, Öğrenim Durumu. The table contains 20 rows of data. On the right, a dialog box titled 'Uyarı' (Warning) is open, asking 'Verileri değiştirmek istediğinize emin misiniz?' (Are you sure you want to change the data?). Below the dialog, there is a 'Medeni Durum' dropdown menu with a list of values: EVLI, BEKAR, DUL, BOŞANMIŞ. To the right of this list, there are buttons for 'Veri Ekle', 'Veri Çıkar', 'Veri Güncelle', and 'Değişken Değerlerini Değiştir'. A checkbox labeled 'Değerler Değiştirilecek mi?' is checked.

3.1.5 | Veri setine apriori algoritmasının uygulanması ve kuralların oluşturulması

Apriori algoritmasında birliktelik kuralı oluşturulurken kullanılan minimum bir destek sayısı ile asgari güven değerlerinin belirlenmesi ve yazılımda ilgili alana tanımlanması gerekmektedir. Destek ve güven değerlerinin belirlenmesi, oluşturulacak birliktelik kurallarının ilginçliği açısından önem taşımaktadır. Bu değerlerin düşük girilmesi durumunda birliktelik kuralı sayısı yüksek

çıkacak, bu nedenle değersiz çok sayıda kural ile karşı karşıya kalınacaktır. Düşük girilmesi durumunda ise bu sayı yetersiz olabileceğinden, anlamlı birliktelik kural tahmini yapmak zorlaşacaktır. Bu durumun önüne geçebilmek için farklı düzeyde destek ve güven değerleri uygulanarak sorun çözülebilmektedir. Uygulama yazılımında eşik ve güven değeri alanlarına giriş yapılarak birliktelik kuralları oluşturulur. Destek ve güven değerleri tanımlanarak birliktelik kurallarının oluşturulması Şekil 6' da gösterilmiştir.

Şekil 6. Destek ve Güven Değerleri Girilmesi ile Oluşturulan Birliktelik Kuralları

The screenshot shows the results of the Apriori algorithm. On the left, a list of rules is displayed, each with its support and confidence values. For example, '[21] -> [11]: Destek: 26, Güven: 0.456140350877193'. On the right, a list of generated rules is shown, such as 'Medeni Durum , EVLI , 11' and 'Medeni Durum , BEKAR , 12'. Below the list, there are input fields for 'Eşik Değeri' (0.2) and 'Güven Değeri' (4). Buttons for 'Apriori', 'Data Gridview Aktar', and 'Excel Dosyasına Aktar' are visible. At the bottom, there are buttons for 'Sonuçları Kopyala' and 'Sonuçları Kopyala'.

3.1.6 | Sonuç Çıktısının Oluşturulması

Birliktelik kuralları oluşturulduktan sonra bir tablo nesnesine aktarılır. Bu kurallar tablo nesnesinden bir

veri tabanına aktarılabilen veya seçilen bölümü kopyalanarak bir metin belgesi olarak saklanabilmektedir. Oluşturulan birliktelik kurallarının bir çıktı örneği Şekil 7' de gösterilmiştir.

Şekil 7. Birliktelik Kurallarının Çıktı Görüntüsü

İşsiz; Okur Yazar Olmayan	> Kpss Puanı Yok	> Güven: 0,993 Destek: 134 Lift: 1,286
Okur Yazar Olmayan	> Kpss Puanı Yok; İşsiz	> Güven: 0,993 Destek: 134 Lift: 1,723
Okur Yazar Olmayan	> Normal; Yabancı Dil(Hayır); Kpss Puanı Yok	> Güven: 0,993 Destek: 134 Lift: 2,034
Yabancı Dil(Hayır); Okur Yazar Olmayan	> Normal; Kpss Puanı Yok	> Güven: 0,993 Destek: 134 Lift: 1,301
Normal; Okur Yazar Olmayan	> Yabancı Dil(Hayır); Kpss Puanı Yok	> Güven: 0,993 Destek: 134 Lift: 2,017
Okur Yazar Olmayan	> Normal; Kpss Puanı Yok; İşsiz	> Güven: 0,993 Destek: 134 Lift: 1,732
İşsiz; Okur Yazar Olmayan	> Normal; Kpss Puanı Yok	> Güven: 0,993 Destek: 134 Lift: 1,301
Normal; Okur Yazar Olmayan	> Kpss Puanı Yok; İşsiz	> Güven: 0,993 Destek: 134 Lift: 1,723
Yabancı Dil(Hayır); İşsiz; Okur Yazar Olmayan	> Kpss Puanı Yok	> Güven: 0,993 Destek: 134 Lift: 1,286
Okur Yazar Olmayan	> Yabancı Dil(Hayır); Kpss Puanı Yok; İşsiz	> Güven: 0,993 Destek: 134 Lift: 2,612
İşsiz; Okur Yazar Olmayan	> Yabancı Dil(Hayır); Kpss Puanı Yok	> Güven: 0,993 Destek: 134 Lift: 2,017
Yabancı Dil(Hayır); Okur Yazar Olmayan	> Kpss Puanı Yok; İşsiz	> Güven: 0,993 Destek: 134 Lift: 1,723
Okur Yazar Olmayan	> Normal; Yabancı Dil(Hayır); Kpss Puanı Yok; İşsiz	> Güven: 0,993 Destek: 134 Lift: 2,626
Yabancı Dil(Hayır); İşsiz; Okur Yazar Olmayan	> Normal; Kpss Puanı Yok	> Güven: 0,993 Destek: 134 Lift: 1,301
Normal; Okur Yazar Olmayan	> Yabancı Dil(Hayır); Kpss Puanı Yok; İşsiz	> Güven: 0,993 Destek: 134 Lift: 2,612
Yabancı Dil(Hayır); Okur Yazar Olmayan	> Normal; Kpss Puanı Yok; İşsiz	> Güven: 0,993 Destek: 134 Lift: 1,732
İşsiz; Okur Yazar Olmayan	> Normal; Yabancı Dil(Hayır); Kpss Puanı Yok	> Güven: 0,993 Destek: 134 Lift: 2,034
Daha iyi şartlarda iş arayan; İlköğretim	> Normal; Yabancı Dil(Hayır)	> Güven: 0,991 Destek: 111 Lift: 1,757
Daha iyi şartlarda iş arayan; İlköğretim; Kpss Puanı Yok	> Normal; Yabancı Dil(Hayır)	> Güven: 0,991 Destek: 109 Lift: 1,757
İlköğretim	> Normal; Yabancı Dil(Hayır)	> Güven: 0,988 Destek: 165 Lift: 1,752
İlköğretim; Kpss Puanı Yok	> Normal; Yabancı Dil(Hayır)	> Güven: 0,988 Destek: 162 Lift: 1,751
Yabancı Dil(Hayır); İlk kez iş hayatına atılan	> İşsiz	> Güven: 0,985 Destek: 137 Lift: 1,303
Yabancı Dil(Hayır); İlk kez iş hayatına atılan	> Normal; İşsiz	> Güven: 0,985 Destek: 137 Lift: 1,310
Daha iyi şartlarda iş arayan; İlköğretim	> Kpss Puanı Yok	> Güven: 0,982 Destek: 111 Lift: 1,272
Daha iyi şartlarda iş arayan; İlköğretim	> Yabancı Dil(Hayır); Kpss Puanı Yok	> Güven: 0,982 Destek: 111 Lift: 1,996
İlköğretim	> Kpss Puanı Yok	> Güven: 0,982 Destek: 165 Lift: 1,272
İlköğretim	> Yabancı Dil(Hayır); Kpss Puanı Yok	> Güven: 0,982 Destek: 165 Lift: 1,996

4 | SONUÇ ve ÖNERİLER

Çalışma sonucunda toplam 555 birliktelik kuralı elde edilmiştir. Sonuç olarak çıkan birliktelik kurallarının tamamı bir anlam taşımamakta, ancak değerlendirme sonucu anlamlı sonuçlara ulaşmak mümkün olmaktadır. Algoritmanın birliktelik kurallarını hesaplama süresi güven ve destek değerleri ile orantılı olarak azalma ya da artış göstermektedir.

Oluşturulan birliktelik kuralları yorumlanırken destek, güven ve ilginçlik değerleri kullanılmaktadır. Örneğin Şekil 7'de ilk sırada bulunan kayıt için destek değerine göre, okur yazar olmayanların %99,3'ünün yabancı dili ve Kamu Personeli Seçme Sınavı (KPSS) puanı yok ve işsiz şeklinde yorumlanmıştır. Lift değerinin 2,612 olması okur yazar olmayan başvuru sahiplerinin yabancı dil ve KPSS puanlarının olmadığı aynı zamanda işsiz oldukları parametreleri arasında kuvvetli bir korelasyon olduğunu göstermektedir.

Oluşan birliktelik kurallarının öneme ve ilginçliğe sahip olduğu düşünülenleri şu şekilde özetlemek mümkündür. İlk kez iş hayatına atılan başvuru sahiplerinin %80'i KPSS puanına sahiptir. İlk kez iş hayatına atılan kadınlar toplam başvurunun %11,86'sını oluşturmakta ve %90'ı işsizdir. Başvuru yapanlar arasında iş sahibi olmayan kadınların oranının %32,04 olduğu tespit edilmiştir.

Eğitim durumları göz önüne alındığında, evli ve ilköğretim mezunu başvuru yapanların tamamı KPSS puanına sahiptir. Genel başvuruların %17'lik kısmı, lisans mezunu işsizlerden oluşmaktadır. Aynı zamanda iş sahibi olan lisans mezunlarının %80'i aktif olarak iş aramaktadır. Ön lisans mezunu başvuru yapan aday oranı ise %11 düzeyindedir. İlköğretim

mezunu KPSS'ye girmemiş kadın adayların %80'i daha iyi şartlarda iş aramaktadır.

Bölgeler bazında bakıldığında en güçlü desteğin Marmara bölgesinde bulunduğu ve bu bölgeden başvuru yapanların %90 oranında KPSS puanına sahip olduğu görülmüştür. Bunun yanı sıra profili olmayan ya da düzenlememiş olan adayların %90'ı KPSS puanına sahiptir.

Engeli ya da hüküm durumu olmayan normal nitelikteki erkek bireylerin %80'i iş sahibi değildir ve aktif olarak iş aramaktadır. İş arayan ancak profilini doldurmamış ya da eksik doldurmuş bireylerin oranı %12,64 seviyesindedir.

Elde edilen sonuçlara göre, İŞKUR özelinde yapılan çalışmanın sonucunda veri madenciliği analizlerinden anlamlı sonuçlar ortaya çıkartılması ve bu sonuçların başvuru düzenlemede kullanılabilmesi mümkün görülmektedir. Böyle bir çalışma yapılmasının çeşitli avantajları vardır. İŞKUR'un elinde bulunan veriden bilgi keşfi yapılması neticesinde, tahminlerin ve durum değerlendirmesinin gerçekleştirilmesi, ayrıca geleceğe yönelik kararların alınması sağlanabilir.

Bulunan birliktelik kuralları incelendiğinde görülen diğer bir kural, ilk kez iş hayatına atılan kadın bireylerin işverenler tarafından tercih edilmemeleri, aynı durumda olan erkek bireylere göre çok daha güçlü bir birlikteliktir. Bununla birlikte başvuru yapan kadın bireylerin işsizlik oranı da dikkate alınması gereken diğer bir sonuçtur. Bu durum Aile, Çalışma ve Sosyal Hizmetler Bakanlığı tarafından değerlendirilebilir ve politikalarını şekillendirmek için kullanılabilir.

Daha iyi şartlarda iş arayan başvuru sahiplerinin oransal yüksekliğinin de dikkate değer nitelikte olduğunu söylenebilir. Bu yüksek oran memnuniyetsizlik göstergesi olarak da kabul edilebilir. Çalışırken işsiz kalan ve aktif durumda iş arayanların yüksek oranı ise ekonomik tahminlerde dikkate alınabilir.

Değerlendirmeye alınması gereken diğer bir sonuç ise aktif olarak iş arayan, KPSS puanına sahip bireylerin profillerini doldurmamış ya da eksik doldurmuş olmalarıdır. Bu sonuçtan adayların başvuru sürecinde eksik bilgilendirdiği ya da İŞKUR başvurularını önemsemedikleri sonucu çıkarılabilir.

Çalışmanın belirli kısıtları vardır. İŞKUR tarafından sağlanan 2017 yılına ait on aylık veriler ile çalışılmış ve başvuru sahipleri tarafından girilen bilgilerin doğru olduğu kabul edilmiştir.

Elde edilen sonuçlar ve yorumlar incelendiğinde yapılan çalışmanın gelecekteki belirli kararları ve politikaları şekillendirebilir nitelikte olduğu söylenebilir. Bu sayede personel arayan iş yeri sahiplerinin uygun personele ulaşma sürecine katkı sağlanabilir.

KAYNAKÇA

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. PROCEEDINGS OF THE 1993 ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, (s. 207-216). WASHINGTON DC.

Agrawal, R., & Shafer, J. (1996). Parallel Mining of Association Rules: Design, Implementation and Experience. IEEE Transactions on Knowledge and Data Engineering, 962-969.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. Proc. of 20th Intl. Conf. on VLDB, (s. 487-499).

Akpınar, H. (2000, Nisan 1). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. İ.Ü işletme Fakültesi Dergisi (29), 1-22.

Das, A., Ng, W., & Woon, Y. (2001). Rapid Association Rule Mining. International Conference on Information, (s. 474-481). Atlanta.

Fayyad, U., Shapiro, G. P., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a unifying framework. KDD'96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, (s. 82-88). Portland, Oregon.

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques 3rd. San Francisco: Morgan Kaufmann Publishers Inc.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. New York: Springer. doi:10.1007/978-0-387-84858-7

Houtsma, M., & Swami, A. (1995). Set-oriented data mining in relational databases. Data & Knowledge Engineering, 245-262.

Kumbhare, T., & Chobe, S. (2014). An Overview of Association Rule Mining Algorithms. (IJCSIT) International Journal of Computer Science and Information Technologies, 927-930.

Linoff, G., & Berry, M. (2011). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Indianapolis, Indiana: Wiley Publishing, Inc.

Salvatore, O., Perego, R., & Silvestri, C. (2004). A new algorithm for gap constrained sequence mining. SAC '04 Proceedings of the 2004 ACM symposium on Applied computing, (s. 540-547).

Srikant, R., & Agrawal, R. (1996). Mining Quantitative Association Rules in Large Relational Tables. ACM SIGMOD International Conference on Management of Data, (s. 1-12). Montreal.

Zaki, M., & Hsiao, C.-J. (tarih yok). CHARM: An Efficient Algorithm for Closed Itemset Mining. Proceedings of the 2002 SIAM International Conference on Data Mining (s. 457-473). içinde doi:10.1137/1.9781611972726.27