

# Site-içi Aramalar ve Apriori Algoritmasıyla Web Sitesi Ziyaretçilerinin İhtiyaç Tespitine Yönelik Bir Örnek Olay İncelemesi

Veli Özcan BUDAK<sup>1\*</sup>, Elif KARTAL<sup>2</sup>, Sevinç GÜLSEÇEN<sup>2</sup>

<sup>1</sup>Bilgi İşlem Daire Başkanlığı, Kırklareli Üniversitesi, Kırklareli, Türkiye

<sup>2</sup>Enformatik Bölümü, İstanbul Üniversitesi, İstanbul, Türkiye

[veliozcanbudak@gmail.com](mailto:veliozcanbudak@gmail.com), [elifk@istanbul.edu.tr](mailto:elifk@istanbul.edu.tr), [gulsecen@istanbul.edu.tr](mailto:gulsecen@istanbul.edu.tr)

(Geliş/Received: 20.02.2018; Kabul/Accepted: 07.04.2018)

DOI: 10.17671/gazibtd.397142

**Özet**— Bugün web siteleri, milyarlarca insan tarafından kullanılmaktadır. Bir kişinin/topluluğun/kurumun veya markanın; web sitesi aracılığı ile daha çok kişiye ulaşabilmesinde, takipçileri tarafından kabul görmesinde ve böylelikle hedeflediği başarıyı yakalayabilmesinde, site ziyaretçilerinin ihtiyaçlarının tespit edilebilmesi son derece önemlidir. Tespit edilen bu ihtiyaçlar, bir web sitesinin tasarım ve içerik yönünden geliştirilmesinde kilit rol oynamaktadır. Bu çalışmanın amacı; site-içi aramalar ve apriori algoritması kullanılarak web sitesi ziyaretçilerinin ihtiyaç tespitine yönelik bir örnek olay incelemesi sunmaktır. Bu kapsamda, veri seti olarak Kırklareli Üniversitesi web sitesinden ([www.klu.edu.tr](http://www.klu.edu.tr)) elde edilen bir aylık web günlük dosyası kullanılmıştır. Analiz süreci, Veri Madenciliği için Çapraz Endüstri Standard Süreç Modeli (CRISP-DM: CRoss-Industry Standard Process for Data Mining) çerçevesinde ele alınmıştır. Üniversiteye kayıtların başlangıcına denk gelen Ağustos ayına yönelik analizler sonucunda; ziyaretçiler tarafından gerçekleştirilen site-içi aramalarda çoğunlukla “Yatay Geçiş”, “Kayıt Yenileme”, “Ders Programı”, “Ders Kayıtları”, “Harç”, “Kontenjanlar” ve “Akademik Takvim” kelime/kelime gruplarının olduğu gözlemlenmiştir. Bununla birlikte, sırasıyla “Yatay Geçiş” (1177 kez), “Harç” (889 kez) ve “Ders Programı” (600 kez) aramalarının yapılan tüm aramaların başında geldiği tespit edilmiştir. Bu noktadan hareketle, hem Kırklareli Üniversitesi hem de diğer üniversiteler için; “Kayıt Yenileme”, “Harç” ve “Ders Kayıt” aramalarıyla ilgili içeriklerin aynı anda tek bir sayfadan sunulması önerisi bu çalışmanın en belirgin sonuçlarından biridir. Buna ilaveten, “Yatay Geçiş” ve “Kontenjan” aramalarıyla ilgili içeriklerin tek bir sayfadan sunulması önerilmektedir. Aynı zamanda, “Akademik Takvim” aramasına yönelik ilgili içeriğe yönlendiren bir bağlantıyı içeren küçük bir resmin, web sitesi ana sayfasının üst bölümünde, ziyaretçilerin dikkatini çekici uygun bir noktaya yerleştirilmesi önerilmektedir. Çalışma bulguları ışığında geliştirilen diğer öneriler, dönemsel ve sürekli öneriler başlıkları altında sonuç bölümünde paylaşılmıştır.

**Anahtar Kelimeler**— Apriori Algoritması, Web Günlük Dosyası Analizi, Web Madenciliği, Ziyaretçi İhtiyaçları

## A Case Study on Identifying Visitor Needs of a Website with In Site Search and Apriori Algorithm

**Abstract**— Nowadays, websites are used by billions of people. The identification of the visitor needs of website is very important for a(n) person/community/organization or a brand to reach more people through a website, to be accepted by website followers, and so to achieve targeted success. These identified needs play a key role in improving a website in terms of design and content. The aim of this study is to provide a case study to identify visitor needs of a website by using in site search and the apriori algorithm. In this context, a monthly web log file which is obtained from Kırklareli University website ([www.klu.edu.tr](http://www.klu.edu.tr)) was used as data set. Analysis process is discussed in the context of CRISP-DM: CRoss-Industry Standard Process for Data Mining. As a result of the analyses made for August which is the beginning of the enrollment to the university, it is observed that word/word groups of “Undergraduate Transfer”, “Re-enrollment”, “Syllabus”, “Course Registration”, “Tuition”, “Quota” and “Academic Calendar” mostly exist in in-site searches performed by visitors. In addition, it is found that the searches of “Undergraduate Transfer” (1177 times), “Tuition” (889 times) and “Syllabus” (600 times) are at the top of the all in site searches list, respectively. In this manner, for both Kırklareli University and other universities, the suggestion that contents related to “Re-enrollment”, “Tuition” and “Course Registration” searches should be presented on the same page at once is one of the most significant results of this

study. Also, it is recommended that contents related to “Undergraduate Transfer” and “Quota” searches should be presented on the same page at once, too. Moreover, it is suggested that a small image which has a link to relevant content for the "Academic Calendar" search should be placed to a suitable point at the top of the website homepage where is attractive for visitors. Other suggestions developed in the light of study findings are shared in the conclusion section under the headings of periodic and continuous suggestions.

**Keywords**— Apriori Algorithm, Web Log Analysis, Web Mining, Visitor Needs

## 1. GİRİŞ (INTRODUCTION)

Web, milyarlarca insan tarafından kullanılan ve barındırdığı veri miktarı sürekli büyüyen bir teknolojidir. Sadece beş yıl öncesi ile günümüzdeki web sitesi sayısı karşılaştırılırsa; 2013 yılında yaklaşık 670 milyon olan web sitesi sayısının günümüzde 1 milyar 350 milyona yaklaştığı görülebilir [1]. Dijital okuryazarlığın ve internet erişiminin artması, artık neredeyse tüm işlemlerin elektronik ortamda kolaylıkla gerçekleştirilebilmesi, web sitesi sayısındaki bu artışın başlıca nedenleri arasında gösterilebilir. Bir web sitesi, bir bireyin/topluluğun/kurumun ya da markanın elektronik ortamdaki yüzü olarak da değerlendirilebilir. Bu nedenle, site ziyaretçilerinin sitede aradığını kolaylıkla bulabilmesi, istediği bilgiye kolaylıkla erişebilmesi gerekmektedir. Kullanıcıların aradıkları bilgiye kolayca erişmelerini sağlamak ise sistem yöneticilerinin öncelikleri arasında yer almaktadır [2]. Graham [3], bir sistemi kullanan kişilerin farklı arka planlara ve bilgi seviyelerine sahip olduklarını belirtmiştir. Dolayısıyla, kullanıcı farklılıklarına hızlı bir şekilde cevap verebilecek geliştirmelerin bir web sitesinde de yapılmasının önemli olduğu rahatlıkla söylenebilir. Bu da site yöneticisinin, site ziyaretçilerinin ihtiyaçlarını doğru bir şekilde tespit etmesi ve bu doğrultuda site tasarımı ve içeriğinde geliştirme ve iyileştirmeler yapmasıyla mümkündür. Web madenciliği, bu konuda araştırmacılara fayda sağlamaktadır.

Temelde web madenciliği, veri madenciliğinin webe odaklanmış bir alt çalışma alanıdır. Veri madenciliği büyük ölçekli veri içinde gizli değerli bilginin keşfidir [4]. Büyük hacimdeki ve farklı tiplerdeki veriden anlamlı kalıplar oluşturmayı ve bu veri arasındaki ilişkilerin keşfedilerek bilginin elde edilmesini amaçlayan bir alandır [5]. Web madenciliği ise, web üzerinden kullanışlı bilginin çıkarılması ve tekrar eden şablonların tespit edilmesi sürecidir [6]. Web madenciliği sayesinde, kullanıcılarla ilgili çıkarımlar yapılabilmekte ve kullanıcıların eğilimlerine göre içerik düzenlenebilmektedir [7]. Büyük veri içinde gizli kalmış ve işe yarar bilginin elde edilmesi için web ortamına özgü veri kullanılmaktadır. Web günlük (log) dosyaları, webdeki veri madenciliği analizlerinde kullanılan başlıca veri kaynaklarından biridir. Web günlük dosyaları, web sunucularında bulunan ve sistem üzerinde gerçekleşen eylemlerin tutulduğu dosyalardır [8]. Ziyaretçilerin, elektronik ortamda gerçekleştirdiği her türlü işlemin izi bu dosyalarda tutulmaktadır. İlk kullanım amaçları sunucu yöneticilerinin sunucu ağ genişliği ve

kapasitesinin izini sürmeleri olmasına rağmen [9], bu dosyaların analizleri sayesinde, çoğunlukla hizmete açılmış ürünlerin işleyişi hakkında ve ziyaret edilen sayfalar, ziyaretçi miktarındaki azalmalar, trafik dalgalanmaları gibi web sitesi hakkında birçok veri sağlanmaktadır [10]. Farklı koşullar altındaki çok sayıda kullanıcıdan kolaylıkla elde edilen bu veri [7] sayesinde, bir web sitesinde yer alan ürünlerin sürekliliğinin garanti altına alınması, gözle anında fark edilemeyecek eksikliklerin giderilmesi, vb. geliştirmeler rahatlıkla gerçekleştirilebilir. Web madenciliği kendi içinde; web içerik madenciliği (web content mining), web yapı madenciliği (web structure mining) ve web kullanım madenciliği (web usage mining) olmak üzere üç gruba ayrılmaktadır [6, 8, 12-14]. Web içerik madenciliği, temel olarak yazı, resim, ses, video, meta veri ve köprü gibi çeşitli türlerdeki veriyi içeren web içeriklerinden kullanışlı bilginin çıkarılması olarak tanımlanmaktadır [15]. Web yapı madenciliği, web bağlantı yapılarındaki modellerin keşfedilme işlemidir [15]. Web kullanım madenciliği ise, web tabanlı uygulamaların ihtiyaçlarını anlamak ve daha iyi hizmet sunmak amacıyla web verisinden kullanım desenlerinin/örüntülerinin keşfedilmesi için veri madenciliği tekniklerinin uygulanmasıdır [16]. Web kullanım madenciliğinin temel hedefi; kullanıcılar web ile etkileşim halindeyken, kullanıcıların tarama davranışlarının anlaşılması, resmedilmesi ve modellenmesidir [6]. Bu bilgiler doğrultusunda, bu çalışmada web kullanım madenciliğinden faydalanılarak web sitesi ziyaretçilerinin ihtiyaç tespitine yönelik bir örnek olay incelemesi gerçekleştirilmiştir. Bu kapsamda yazarlar, bir sonraki bölümde çalışmanın arkaplanını yansıtabilmek için literatürde gerçekleştirilen benzer çalışmalara ve bu çalışmaların dikkat çekici sonuçlarına yer vermiştir. Sonrasında ziyaretçilerin doğrudan site-içi aramalarda kullandığı anahtar kelime ve kelime gruplarının web günlük dosyalarındaki bağlantı adreslerinden süzülerek, apriori algoritmasıyla analiz edildiği bir örnek olay incelemesine ilişkin detaylara yer verilmiştir.

## 2. LİTERATÜR TARAMASI (LITERATURE REVIEW)

Das ve Türkoğlu [17] tarafından yeni bir web günlük dosyası analizi yaklaşımı getirmek amacıyla yapılan çalışmada, web günlük dosyalarından daha başarılı sonuçların elde edilmesini sağlayacak bir metod önerilmiştir. Önerilen bu metotta, bir üniversiteden elde

edilen web günlük dosyası kullanılmıştır. Web sitesinde yer alan bağlantı adresleri arasındaki ilişkilerin tespit edildiği çalışma sonucunda, web madenciliğinin web sitelerinde yer alan yönlendirme/menü yapılarındaki değişim konusunda karar desteği sağladığı belirtilmiştir. Bununla birlikte; kullanıcı davranışlarının analiz edilmesiyle web sitesi ziyaret oranlarının artırılabilmesi, web sitelerindeki bağlantı adresleri arasındaki ilişkilerde ortaya çıkarılan çeşitli davranış kalıplarının analiziyle de, web sitelerinin hem daha iyi hizmet verebileceği hem de verimliliğinin artırılabilmesi sonuçlarına ulaşılmıştır.

Tyagi, Solanki ve Wadhwa [18], bilgisayarlar için yedekleme ve senkronizasyon hizmetinin sunulduğu bir web sitesinin web günlük dosyası analizini gerçekleştirmiştir. Ziyaretçilerin karşılaştığı sistem hatalarının tespiti ve bu tespitlerden yola çıkılarak web sitesinin etkililiğinin artırılmasına yönelik bulguların ortaya çıkarılması bu çalışmanın amacını oluşturmuştur. Bir haftalık web günlük dosyası verisi üzerinden yapılan analizler sonucunda web sitesinde yaşanan hatalara yönelik sonuçlara, sayfa görüntüleme, ziyaretçi miktarı ve ziyaretlerde kullanılan bant genişliği istatistiklerine ulaşılmıştır. Yazarlar, çalışma sonucunda ortaya çıkarılan istatistiklerin ve çalışmalarındaki benzer şekilde gerçekleştirilecek analizlerin farklı web sitelerinin etkililiğinin artırılması için kullanılabilmesini belirtmişlerdir. Suneetha ve Krishnamoorthi [19], kullanıcıların bireysel ihtiyaçlarının belirlenmesi amacıyla C4.5 karar ağacı algoritmasını kullanarak NASA web sunucusundan elde edilen web günlük dosyaları üzerinden analizler gerçekleştirmiştir. Analiz sonuçları sayesinde satın alma eğiliminde olan sürekli kullanıcıların erişim şablonları ortaya çıkarılmıştır. Çalışma yönteminin, e-ticarette ilgili sistemlerde hem var olan kullanıcıların kaybedilmemesinde, hem de yeni kullanıcıların sisteme kazandırılmasında önemli katkılar sağlayacağı belirtilmiştir. Aynı zamanda, analizlere yönelik zaman ve sistem hafızasının daha verimli kullanıldığı ortaya konmuştur.

Carmona ve diğerleri [20], zeytinyağı satışının gerçekleştirildiği bir web sitesini ziyaret eden kullanıcıların, bu siteyle olan etkileşimleri hakkında bilgileri ortaya çıkarmak amacıyla web kullanım madenciliğinden yararlanmıştır. Çalışmada analiz edilen veri, Google Analytics web günlük dosyaları üzerinden elde edilmiştir. Çeşitli algoritmaların bir arada kullanıldığı bu çalışma sonucunda, web sitesine erişim sağlayan kullanıcıların %10'luk bir bölümünün Internet Explorer (IE) tarayıcısını kullandığı tespit edilmiş ve sistem yöneticilerinin web sitesinde yer alan görselleri IE ve benzeri tarayıcılara yönelik olarak geliştirmeleri gerektiği vurgulanmıştır. Bununla birlikte, web sitesinde yer alan sayfaların düşük ziyaret oranlarına sahip olduğu için referans web sitelerine dikkat edilmesi gerektiği belirlenmiştir. Santra ve Jayasudha [21], sürekli kullanıcıların davranışları üzerinden C4.5 ve basit (naive) bayes gibi iki farklı veri madenciliği algoritmasıyla web günlük dosyası analizi gerçekleştirmiştir. Çalışmada, daha

az bellekle daha hızlı analiz gerçekleştirebilmek için hangi algoritmanın daha başarılı olduğunun tespit edilmesi amaçlanmıştır. Çalışma sonucunda basit bayes algoritmasının üstünlüğü vurgulanmıştır. Bununla birlikte algoritmanın, e-ticaret uygulamalarında web sayfalarının kişiselleştirilmesi, tavsiye edilmesi gibi amaçlarla web günlük dosyası analizi gerçekleştirilirken kullanılabilmesi ifade edilmiştir. Sujatha ve Punithavalli [22], web günlük dosyaları üzerinden kullanıcıların gezinme davranışlarını tahmin edebilen bir sistem geliştirmişlerdir. Bu sistem, gereksiz web günlük dosyası satır kayıtlarını temizleme, potansiyel kullanıcıların tespiti, kullanıcıların gezinme davranışı keşfi ve web sitesine gelebilecek istek tahmini olmak üzere toplam dört aşamadan oluşmaktadır. Web günlük dosyalarının belirtilen aşamalardan geçirilerek analiz edildiği bu çalışma sonunda, çalışmada kullanılan tekniğin kullanıcıların gezinme davranışlarının sınıflandırılmasında etkili olduğu belirtilmiştir.

Lopes ve Roy [23], bir e-ticaret web sitesi kullanıcılarına, kullanıcıların üyelik durumlarına bakılmaksızın gerçek zamanlı dinamik öneriler sunan bir sistem geliştirmiştir. Geleneksel öneri sistemleriyle karşılaştırıldığında daha başarılı sonuçların elde edildiği bu çalışmada web kullanım madenciliğinden yararlanılmıştır. Sonuçlar, sistem sayesinde üye kullanıcılar üzerinde yaklaşık %80-%85, üye olmayan kullanıcılarda ise %65-%70 arası verimliliğin artırılmasına katkı sağlandığını göstermiştir. Sistemin var olan kullanıcıların korunması ve yeni kullanıcıların dikkatinin çekilmesi noktasında potansiyele sahip olduğu belirtilmiştir. Shih ve Huang [24], bir web sitesi üzerinde kullanıcılardan istenen bilgileri web kullanım madenciliği ile sınıflandırarak, benzer kullanıcıları gruplamaya yönelik bir çalışma gerçekleştirmişlerdir. K-ortalamlar ve apriori algoritmaları kullanılarak yapılan analizler sonucunda yedi farklı kullanıcı grubu ortaya çıkarılmıştır. Ortaya çıkarılan bu grupların her birinin ihtiyaçlarına yönelik olarak hangi alanlarda bilgiye erişmek istedikleri tespit edilmiştir. Bu sayede, her bir gruba özel pazarlama stratejilerinin yapılabileceği belirtilmiştir. Chauhan ve Tarar [6], kullanıcıların gezinme davranışlarının tahmin edilmesini amacıyla web kullanım madenciliğiyle web günlük dosyası analizi gerçekleştirmiştir. Web günlük dosyaları, 29 şehirde yer alan müşteri ilişkileri yönetim sisteminde kayıtlı müşteriler üzerinden elde edilmiştir. Çalışma sonucunda kullanıcılara daha hızlı dönüş yapabilmek için davranış tahmini yapmanın önemi vurgulanmış ve çalışmada izlenen yöntemin hem daha doğru sonuçlar ortaya çıkardığına hem de daha az hata oranıyla etkili olduğuna değinilmiştir. Zhang ve diğerleri [25], matris apriori algoritması kullanarak iki farklı web sitesinde yapılan aramaları birliktelik kuralı/kuralları bulmak amacıyla web günlük dosyaları üzerinden analiz etmiştir. Web sitelerinde gerçekleştirilen arama kelimelerinden hareketle site ziyaretçilerine, "Yaşam ansiklopedisi arayan kullanıcılara popüler bilgi ve siyasetle ilgili içerikler gösterilebilir" ya da "Spor ayakkabı arayan kullanıcılara günlük ayakkabı, kanvas ayakkabı ve sandalet içeren içerikler gösterilebilir" gibi öneriler getirilmiştir. Çalışma sonunda, kullanılan analiz yönteminin kullanıcı

davranışları ile kullanıcıların gerçekleştirdiği işlemler arasındaki ilişkiyi ortaya çıkarmada başarılı olduğu tespit edilmiştir.

### 3. YÖNTEM (METHOD)

Web sitesi ziyaretçilerinin ihtiyaç tespitine yönelik gerçekleştirilen bu örnek olay incelemesinde, Veri Madenciliği için Çapraz Endüstri Standard Süreç Modeli (CRISP-DM: Cross-Industry Standard Process for Data Mining) [26] temel alınmıştır. Süreç modeli; problemin tanımlanması, veriyi anlama, veri hazırlama, model kurma, model değerlendirme ve seçimi ile modelin uygulamaya geçirilmesi gibi altı adımdan oluşmaktadır [27]. Sürece ait ilk dört başlığa bu bölümde yer verilmiş olup, model değerlendirme ve seçimi aşaması çalışmanın Bulgular bölümünde, modelin uygulamaya geçirilmesi aşaması ise Tartışma ve Sonuç bölümlerinde paylaşılmıştır.

#### 3.1. Problemin Tanımlanması (Business Understanding)

Bu çalışmada ele alınan temel problem, “web madenciliği ile web sitesi ziyaretçilerinin ihtiyaçlarının tespit edilebilmesi”dir. Bu sebeple bu çalışmanın amacı, “site-içi aramalar ve apriori algoritması kullanılarak web sitesi ziyaretçilerinin ihtiyaç tespitine yönelik bir örnek olay incelemesi sunmak” olarak belirlenmiştir. Bu kapsamda, veri seti olarak Kırklareli Üniversitesi (KLU) web sitesinden (www.klu.edu.tr) elde edilen bir aylık web günlük dosyası kullanılmıştır. Örnek olay incelemesi için odaklanılan Kırklareli Üniversitesi; öğrencisi ve personeliyle yaklaşık 30.000 kişilik bir mevcuda sahiptir. Bununla birlikte, KLU bünyesinde yer almayan; ancak KLU ile ilgili gelişmeleri takip eden diğer bireyler de düşünüldüğünde, gelişmekte olan bu üniversitenin gittikçe daha fazla kişinin ilgilendiği bir kurum haline geleceği söylenebilir. Bu bilgiler ışığında, üniversite vitrini olan

web sitesinin ziyaretçi ihtiyaçlarını karşılaması noktasında sürekli gelişmeye ihtiyaç duyacağı aşikârdır. Çünkü, başvurularla ilgili bilgi edinme ve kayıt dönemi gibi öğrencilerin üniversitedeki gelişmelerden haberdar olmak için sıklıkla başvurdukları adreslerin başında KLU web sitesi gelmektedir. Bu nedenle; bu çalışmanın giriş ve literatür taraması bölümlerinde ortaya konduğu, amacında da vurgulandığı biçimde, KLU web sitesini ziyaret eden kullanıcıların ihtiyaçlarının tespit edilmesi ve web sitesinin tasarım ve içerik anlamında bu ihtiyaçları giderecek biçimde yenilenmesi büyük önem taşımaktadır. Bu nedenle bu çalışma sonuçları hem KLU web sitesi ziyaretçileri hem de üniversite açısından fayda sağlayacaktır. Ayrıca, site-içi aramaların ve apriori algoritmasının kullanıldığı bu örnek olay incelemesinin, bu alanda gerçekleştirilebilecek yeni çalışmalara ışık tutacağına inanılmaktadır.

#### 3.2. Veriyi Anlama (Data Understanding)

Bu çalışmada, KLU web sitesinin 2017 yılının Ağustos ayına ait web günlük dosyaları analiz edilmiştir. Web günlük dosyası analizlerinin bu çalışmada kullanılmasının başlıca sebebi, literatür taraması bölümünde verilen çalışmalarda da görülebileceği gibi hedef ziyaretçi ihtiyaçlarının belirlenmesinde oldukça etkin olmasıdır. Bununla birlikte, KLU web sitesi yaklaşık on yılı aşkın süredir aktif bir web sitesidir ve web günlük dosyası analizleri, hâlihazırda işleyen sistemin başarısını sistemin işleyişini bozmadan analiz edebilme fırsatını sunmaktadır.

Web günlük dosyası analizinde kullanılan tarih aralığı belirlenirken web sitesinde yaşanan yoğunluk bilgisi temel alınmıştır. Web sitesinde yoğunluğun tespit edilmesi amacıyla son yıla ilişkin KLU Google Analytics kullanıcı miktarı istatistiklerinden yararlanılmıştır (Şekil 1). Google Analytics, kullanıcı davranışlarının takip edilmesini ve analizini sağlayan popüler web analitiği araçlarından biridir [28]. Şekil 1’de grafiğin 1 Ocak - 31 Aralık 2017 tarihleri arasında Ağustos ayında tepe noktasına ulaştığı görülebilir.



Şekil 1. KLU Google Analytics kullanıcı miktarı istatistikleri – 2017  
(User counts statistics of KLU Google Analytics - 2017)

Analizlerin gerçekleştirildiği veri setini oluşturan web günlük dosyalarının özelliklerine Tablo 1’de yer verilmiştir.

Tablo 1. Analiz edilen KLU web günlük dosyaları ve özellikleri - Ağustos 2017  
(Analyzed KLU log files and its specifications – August 2017)

Web Günlük Dosyası Numarası	Başlangıç Tarihi	Bitiş Tarihi	Satır Miktarı	Dosya Boyutu (MB)
1	1.Ağu	6.Ağu	2.606.132	766,40
2	6.Ağu	13.Ağu	7.811.186	2.374,05
3	13.Ağu	20.Ağu	6.450.716	1.958,06
4	20.Ağu	27.Ağu	5.642.585	1.701,18
5	27.Ağu	31.Ağu	2.630.354	793,03
<b>Toplam</b>			<b>25.140.973</b>	<b>7.592,72</b>

Veri seti, web sitesinin barındırıldığı sunucuda kurulu olan ve Apache firması tarafından geliştirilmiş olan ücretsiz http sunucu uygulaması üzerinden elde edilmiştir [29]. Web günlük dosyaları, sahip oldukları formatlar bakımından aşağıdaki üç başlıkta sınıflandırılabilir [30]:

- W3C Web Günlük Dosya Formatı
- Microsoft IIS Web Günlük Dosya Formatı
- NCSA Web Günlük Dosya Formatı

KLU web günlük dosyaları NCSA web günlük dosya formatına sahiptir. Bu web günlük dosya formatı; FTP siteleri haricindeki web sitelerinde kullanılan bir format türüdür ve ziyaretçiler hakkında temel bilgileri kayıt altına almaktadır [30]. Bu format türünün daha iyi anlaşılabilmesi için Şekil 2’deki gibi bir web günlük dosya satırı örnek verilebilir:

```
79.123.216.37 - - [01/Aug/2017:00:00:32 +0300]
"GET /rss/Sayfa_Gruplari/74-duyurular/detay/
HTTP/1.0" 200 5274
```

Şekil 2. Örnek bir NCSA web günlük dosyası satırı  
(A sample row of NCSA web log)

Bu türdeki bir web günlük dosyası satırı, aralarında boşluk olacak şekilde farklı alanlara sahiptir ve bu alanların her biri ziyaretçilerden elde edilen farklı türdeki bilgiyi taşır. Şekil 2’deki örnek satır boşluk karakterlerden ayrıldığında Tablo 2’deki bilgilere ulaşmaktadır:

Tablo 2. Örnek bir NCSA web günlük dosyası satırından elde edilen bilgiler  
(Information obtained from a sample row of NCSA web log)

Tür	Erişilen Bilgi
Ziyaretçi IP	79.123.216.37

Tür	Erişilen Bilgi
Ziyaretçi	-
Belirleyicisi	-
Kullanıcı Adı	-
Erişim Zamanı	[01/Aug/2017:00:00:32 +0300]
İstek	"GET /rss/Sayfa_Gruplari/74-duyurular/detay/ HTTP/1.0"
Durum Kodu	200
Gönderilen Bayt	5274

Web günlük dosyası formatları, kurumsal ya da bireysel hizmet veren her türlü uygulama için ihtiyaca göre farklı şekilde ayarlanabilmektedir. KLU bünyesinde tutulmakta olan web günlük dosyası format yapısından örnek bir satır ve bu satırdan elde edilen veri sırasıyla Tablo 3’de aktarılmıştır.

Tablo 3. KLU web günlük dosyasından örnek bir satır ve elde edilen bilgiler  
(A sample row of KLU web log and information obtained from it)

Örnek Satır	Örnek Satırdan Elde Edilen Veri
87.250.233.5 www.klu.edu.tr - [01/Aug/2017:00:01:22 +0300] "GET /Sayfalar/492-05-07-agustos-2014-tarihli-ogretim-eleman-iilani.klu HTTP/1.1" 200 9662 "-" "Mozilla/5.0 (compatible; YandexBot/3.0; +http://yandex.com/bots)"	Ziyaretçi IP: 87.250.233.5 Ziyaret Edilen Adres: www.klu.edu.tr Kullanıcı Adı: - Erişim Zamanı: [01/Aug/2017:00:01:22 +0300] İstek: GET /Sayfalar/492-05-07-agustos-2014-tarihli-ogretim-eleman-iilani.klu HTTP/1.1 Durum Kodu: 200 Gönderilen Bayt: 9662 Gelen Adres: - Kullanıcı Aracı: Mozilla/5.0 (compatible; YandexBot/3.0; +http://yandex.com/bots)

### 3.3. Veri Hazırlama (Data Preparation)

Web günlük dosyalarından oluşan veri setinin ilk aşamada toplamda 25 milyon satırı aşkın veri içerdiği görülmektedir (Tablo 1); ancak analizler öncesinde veri seti ön-işleme sürecine tabi tutularak analizlere hazır hale getirilmiştir. Bu işlem için aşağıdaki adımlar izlenmiş olup, detayları sırayla açıklanmıştır:

#### Site içi aramalara ait satırların elde edilmesi

KLU bünyesinde bulunan tüm birim web sitelerinin işlem kayıtları tek bir web günlük dosyasında tutulmaktadır. Farklı birimlere ait işlem kayıtlarının analiz sürecinden çıkartılması için öncelikle; Tablo 3’de belirtilen “Ziyaret Edilen Adres” niteliğinde “www.klu.edu.tr” adresi bulunmayan satırlar veri setinden çıkarılmıştır. Daha

sonra, ziyaretçilerin site-içi aramalarına odaklanabilmek için Tablo 3'deki "İstek" niteliğinin "GET/Arama\_aracilar/47-tum-sayfalarda-arama?bilgi=aranan\_kelime/kelime\_grubu" halindeki satırları süzölmüştür. Çünkü, "www.klu.edu.tr" adresinde site-içi aramalar gerçekleştirilirken, web günlük dosyasına kaydedilen satıra ait istek her zaman bu ifadeyle oluşturulmaktadır.

#### *Site içinde aranan kelime/kelime gruplarının elde edilmesi*

Her bir satırdaki "İstek" niteliğinde yer alan veri içinden "aranan\_kelime/kelime\_grubu" niteliği elde edilmiştir. Bu nitelik, ziyaretçilerin web sitesi içinde aradıkları içerikler hakkında doğrudan bilgi veren bölümü oluşturmaktadır ve bu kelime/kelime grupları çalışmanın araştırma yönteminde analizler için odaklanılan esas kısımdır.

Bu iki işlem sonucunda, veri seti 39.097 satıra indirgenmiştir. Veri setinde bulunan nitelikler aşağıda listelenmiştir:

- Ziyaretçi IP
- Ziyaret Edilen Adres
- Kullanıcı Adı
- Erişim Zamanı
- İstek
- Aranan Kelime/Kelime Grupları
- Durum Kodu
- Gönderilen Bayt
- Geline Adres
- Kullanıcı Aracı

#### *Site içi aramalardaki benzer kelimelerin/kelime gruplarının birleştirilmesi*

Ziyaretçiler bir kelimeyi, kelimenin tek harfini bile yanlış yazarak aratabildiği için, sitede gerçekleştirilen tek bir aramanın birden fazla varyasyona sahip olabileceği rahatlıkla söylenebilir. Bu nedenle, 39.097 satırlık veri seti OpenRefine uygulaması [31] kullanılarak düzenlenmiştir. OpenRefine, dağınık verinin düzenlenmesi amacıyla kullanılabilen uygulamalardan biridir. Bu çalışmada OpenRefine, site ziyaretçileri tarafından yapılmış olan aramalardaki yazım hatası içeren ya da ortak bir amaca hitap eden benzer kelime/kelime gruplarının belirli kalıplar altında toplanabilmesi amacıyla kullanılmıştır.

#### *Aynı IP'ye sahip kelimelerin/kelime gruplarının tek bir satırda toplanması*

Benzersiz bir IP'ye sahip her bir ziyaretçi aynı anda sadece bir defa site-içi araması yapabilmektedir. Diğer yandan, bir ziyaretçinin sahip olduğu IP değeriyle birden fazla arama yapabilmesi de mümkündür. Bir ziyaretçinin, sahip olduğu belirli bir IP değeriyle yaptığı farklı aramalar, bu ziyaretçinin bir içeriği bulmakta zorlandığını gösterebilir. Bu olasılık göz önüne alınarak aynı IP'ye sahip ziyaretçi

tarafından (eğer yapılmış ise) yapılan tüm aramalardaki kelime/kelime grupları birleştirilmiştir. Diğer bir ifadeyle, web sitesinde tek bir ziyaretçinin yapmış olduğu düşünülen tüm aramalar tek bir satıra toplanmıştır.

Verinin hazırlanması ve analizi R programlama dili ve RStudio Server (versiyon 1.0.153) [32] ile gerçekleştirilmiştir. R, istatistiksel hesaplama ve grafiklerin elde edilmesi için kullanılan ücretsiz bir yazılım ortamıdır [33]. RStudio Server ise, bir web tarayıcı aracılığıyla kullanılabilen, R için bütünsel bir geliştirme ortamıdır [32]. Analizlerde kullanılan R paketleri şu şekildedir: arules [34], foreach [35], doMC [36], plyr [37], urltools [38].

#### *3.4. Model Kurma (Modeling)*

Bu çalışmada, web günlük dosyalarından elde edilen veri setini analiz etmek için Agrawal ve Srikant [39] tarafından geliştirilen apriori algoritmasından yararlanılmıştır. Apriori, çoğunlukla sepet analizinde satın alınan ürünlerin birbiriyle ilişkilerinin ortaya çıkarılması amacıyla kullanılan bir birliktelik kuralı algoritmasıdır. Elde edilen bu kurallar sayesinde, firmaların müşteri satın alma eğilimlerini belirleyerek satış oranlarını arttırması mümkün olabilmektedir. Apriori ve diğer birliktelik kuralı algoritmaları, çalışmanın literatür taraması bölümünde ifade edildiği biçimde web sitelerinin yapısal organizasyonunun geliştirilmesi [17], kullanıcıların erişim kalıplarının çıkarılması [20], kullanıcıların sınıflandırılması [24], web sitelerinin kişiselleştirilmesi [25] gibi amaçlarla sıklıkla tercih edilmektedir. Birliktelik kurallarına ilişkin temel kavramlar kısaca şu şekilde açıklanmaktadır [39-43]:

$I = \{i_1, i_2, \dots, i_m\}$  tüm öğelerden (bu çalışma için ziyaretçiler tarafından yapılan aramalarda kullanılan her bir kelime/kelime grubu: "yatay geçiş", "harç", "kontenjanlar", vb.) oluşan küme olsun.  $T$  gerçekleştirilen her bir hareketi temsil eder ve aynı zamanda  $T \subseteq I$  şeklindeki öğeler kümesidir. Her  $T$  kendine özgü bir hareket numarasına (transaction id/TID) sahiptir (bu çalışmada "Ziyaretçi IP" niteliği TID olarak kullanılmıştır).  $D$  ise tüm hareketlerden oluşan kümeyi göstermektedir.  $X, D$  kümesindeki bir öğe kümesi olmak üzere,  $D$ 'nin  $X$ 'i içeren hareketlerinin tüm hareketlere oranı  $X$ 'in destek (support) değeri olan  $destek(X)$ 'i vermektedir (Denklemler 1). Eğer bu değer uzmanlar tarafından belirlenen minimum destek değerine büyük eşitse,  $X$  sık öğe kümesi (frequent itemset) olarak adlandırılmaktadır.

$$destek(X) = \frac{|X(t)|}{|D|} \quad (1)$$

$X, Y \subset I, X \neq \emptyset, Y \neq \emptyset, X \cap Y = \emptyset$  olan  $X$  ve  $Y$  gibi iki öğe kümesi alındığında, bir birliktelik kuralı  $X \rightarrow Y$  şeklinde ifade edilmektedir. Destek ve güven (confidence) bu kuralların gücünü ortaya koymak için kullanılan iki



temel ölçüdür. Destek,  $X$  ve  $Y$ 'yi içeren hareketlerin tüm hareketlere oranını gösterirken (Denklem 2), güven değeri  $X$  ve  $Y$ 'yi içeren hareketlerin  $X$ 'i içeren hareketlere oranıdır (Denklem 3).

$$\text{destek}(X \rightarrow Y) = \text{destek}(XUY) \quad (2)$$

$$\text{güven}(X \rightarrow Y) = \frac{|XUY(t)|}{|X(t)|} = \frac{\text{destek}(XUY)}{\text{destek}(X)} \quad (3)$$

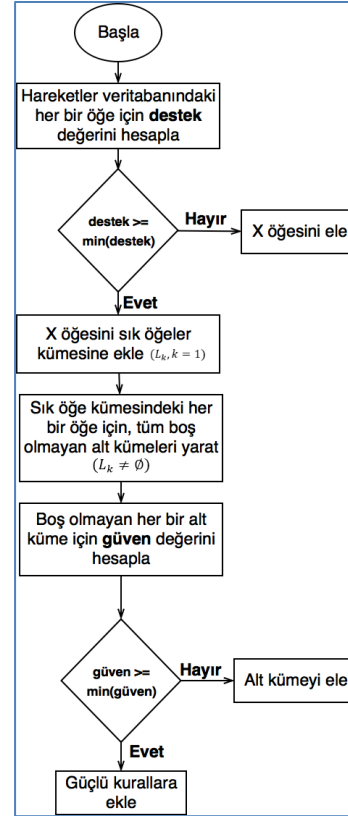
Minimum destek ve güven değerleri kullanıcı tarafından belirlenmekte olup [40], çıkarılan kuralların geçerliliği, aşağıda verilen eşitsizlikleri (Denklem 4, Denklem 5) sağlayıp sağlamamasına bağlıdır [42].

$$\text{destek}(X \rightarrow Y) \geq \min(\text{destek}) \quad (4)$$

$$\text{güven}(X \rightarrow Y) \geq \min(\text{güven}) \quad (5)$$

### 3.4.1. Apriori Algoritması (Apriori Algorithm)

Apriori algoritmasında [41]; öncelikle veri setinde bir öğe içeren tüm sık öğe kümeleri ( $k=1$  olmak üzere  $L_k$ ) bulunur. Ardından  $L_k$ 'dan aday ( $k+1$ )-öge kümeleri elde edilir ( $C_{k+1}$ ). Sadece destek değeri, minimum destek değerine büyük eşit olan aday kümeleriyle  $L_{k+1}$  oluşturulur.  $L_k = \emptyset$  koşulu sağlanana kadar iterasyonlara devam edilir. Algoritmanın akış şeması Şekil 3'te verilmiştir [44, 45].



Şekil 3. Apriori algoritmasının akış diyagramı  
(The flow chart of the apriori algorithm)

Apriori algoritmasıyla gerçekleştirilmiş olan bu çalışmadaki analizlerde, web sitesi ziyaretçilerinin site-içi aramalarda kullandıkları kelime/kelime grupları arasındaki en anlamlı birliktelik kurallarının oluşturulabilmesi için, farklı minimum destek ve güven kriterleri denenmiş; ancak bu denemelerin çoğunda benzer kuralların ortaya çıktığı gözlemlenmiştir. Bu sebeple bu çalışmada, üniversite web sayfasının iyileştirmesine katkı sağlayacağı düşünülen en belirgin kuralların ortaya çıktığı düşünülen minimum destek ve güven değerlerine (Tablo 4) yönelik bulgular bir sonraki bölümde aktarılmıştır.

Tablo 4. Analizlerde kullanılan minimum destek ve güven kriterleri

(Minimum support and confidence criteria which used in analysis)	
Destek	Güven
0,006	0,5
0,006	0,6
0,007	0,2
0,007	0,4
0,008	0,2

### 3.5. Çalışma Sınırlılıkları ve Varsayımları (The Limitations and Assumptions of The Study)

Analizlerde kullanılan veri setinin elde edilmiş olduğu web günlük dosyaları sadece Ağustos 2017'yi kapsamaktadır. Bununla birlikte, ayın ilk gününde arama yapan bir ziyaretçinin, ayın son günü de arama yapmış olma ihtimali

göz önünde bulundurularak aynı IP ile arama yapan ziyaretçilerin aynı kişi olduğu varsayılmıştır.

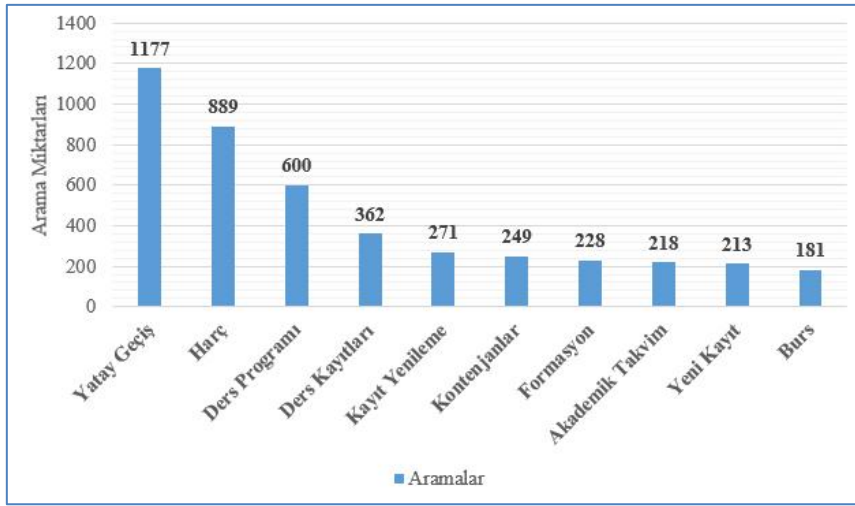
#### 4. BULGULAR (FINDINGS)

Ağustos 2017’de ziyaretçiler tarafından en sık aranan ilk 10 kelime/kelime gruplarına ait grafik Şekil 3’te verilmiştir. Şekil 3’e göre “Yatay Geçiş” (1177 kez), “Harç” (889 kez) ve “Ders Programı” (600 kez) sırasıyla yapılan tüm aramaların başında gelmektedir. Bunları; “Ders Kayıtları”, “Kayıt Yenileme”, “Kontenjanlar”, “Formasyon”, “Akademik Takvim”, “Yeni Kayıt” ve “Burs” izlemektedir.

Apriori algoritması ile Tablo 4’teki minimum destek ve güven kriterleri kullanılarak gerçekleştirilen analizler

sonucunda; Şekil 4’teki kelime/kelime gruplarının yer aldığı birliktelik kuralları da tespit edilmiştir. Aşağıdaki tablolarda sırasıyla Tablo 4’te belirtilen destek ve güven değerlerine göre ortaya çıkmış olan kurallar gösterilmektedir.

Tablo 5’te minimum destek 0,006, minimum güven ise sırasıyla 0,5 ve 0,6 alınarak yapılan analiz sonuçları yer almaktadır. Tablo 5 incelendiğinde, her iki güven kriterinde de benzer kuralların ortaya çıktığı görülmektedir. “Yatay Geçiş Başvuruları”nı arayanların %60’ı “Yatay Geçiş” kelimesini de aratmıştır (Sıra 5). Yine “Ders Programı” ve “Yatay Geçiş” kelimelerini arayanların %60’ı “Harç” kelimesini de aratmıştır (Sıra 7) ve bu üç aramayı bir arada gerçekleştirenler, tüm aramaların %0,6’sını oluşturmaktadır.



Şekil 4. En fazla aranan ilk 10 kelime/kelime grupları sıralaması  
(The top 10 ranking of the most frequently searched word/word group)

Tablo 5. Minimum destek kriteri 0,006 olarak yapılan analizlerdeki oluşan kurallar

(Created rules by the analysis made as the minimum support criterion is 0,006)

Minimum Güven: 0,5				
#	Aranan Kelime/Kelime Grupları	Yönelim	Aranan Kelime/Kelime Grupları	Tekrarlanma Miktarı
1	Yatay Geçiş Başvuruları	→	Yatay Geçiş	49
2	Harç, Yatay Geçiş	→	Ders Programı	31
3	Ders Programı, Harç	→	Yatay Geçiş	31
4	Ders Programı, Yatay Geçiş	→	Harç	31
Minimum Güven: 0,6				
#	Aranan Kelime/Kelime Grupları	Yönelim	Aranan Kelime/Kelime Grupları	Tekrar Miktarı
5	Yatay Geçiş Başvuruları	→	Yatay Geçiş	49
6	Ders Programı, Harç	→	Yatay Geçiş	31
7	Ders Programı, Yatay Geçiş	→	Harç	31

Tablo 6’da minimum destek 0,007, minimum güven ise sırasıyla 0,2 ve 0,4 alınarak yapılan analiz sonuçları yer almaktadır. Tablo 6’nın, Tablo 5’teki bazı kuralları kapsadığı ve yeni kuralları da ortaya koyduğu gözlemlenmiştir. Tablo 6 üzerinde her iki güven kriterinden yola çıkıldığında, “Akademik Takvim” ve “Takvim” aramaları (Sıra 2, 3 ve 8) göze çarpmaktadır. Bu sebeple, “Akademik Takvim” ile ilgili içeriğin daha anlaşılabilir bir şekilde web sitesinden sunulması sağlanabilir. Nielsen [46] tarafından gerçekleştirilen göz izleme çalışmalarında, insanların web sitelerini çoğunlukla F şeklinde tarayarak okuduğunu belirtmiştir. F tarama şekli, ilk adımda kullanıcıların bir sitede en üst sol kısımdan sağa doğru okumaya/taramaya başladığı, ikinci adımda bir alt satıra geçerek yine aynı yönde fakat yatayda daha az içeriği taradığı, üçüncü adımdaysa kullanıcıların dikey olarak aşağıya doğru tarama yapmalarını anlamına gelmektedir [46]. Bu durumda, web sitelerinde bir içerik ne kadar sol en üste yakın olursa o kadar dikkat çekme olasılığının arttırılacağı belirtilebilir. Dolayısıyla, belirtilen geliştirme önerisine yönelik olarak, “Akademik Takvim” başlığının bulunduğu ve ilgili içeriğe yönlenen bir



bağlantıyı içeren küçük bir resmin, web sitesi ana sayfasındaki menünün hemen altında en sola yaslı olacak şekilde sunulmasının bu ihtiyacı karşılayabileceği belirtilebilir.

“Kontenjanlar”ı arayanların %20’si “Yatay Geçiş” kelimesini de aratmıştır (Sıra 6). Bu aramalar Tablo 5’teki kurallarla desteklendiğinde, “Kontenjanlar”, “Yatay Geçiş”, “Harç” ve “Ders Programı” aramalarındaki ihtiyaçları giderebilecek bir web sayfasının hazırlanmasının, ziyaretçilerin ihtiyaçlarını giderebileceği rahatlıkla söylenebilir.

Benzer bir durum Tablo 6 üzerindeki “Kayıt Yenileme” ve “Ders Kayıtları” (Sıra 4 ve 5) aramalarında da görülmektedir. “Kayıt Yenileme”yi arayanların %20’si “Ders Kayıtları”nı (Sıra 5), yine “Kayıt Yenileme”yi arayanların %20’si “Harç” kelimesini de (Sıra 4) aramıştır. Bu noktada, “Kayıt Yenileme”, “Ders Kayıtları” ile “Harç” bilgileri sitede birbirine yakın yerlere alınabilir. “Ders Programı” ve “Yatay Geçiş”i arayanlar (Tablo 5) ile “Kayıt Yenileme” kelimesini arayanların (Tablo 6) “Harç” kelimesini de aradığı ortaya çıkmıştır. Bu nedenle site içindeki revizyon bu üç başlık üzerinden de yapılabilir.

Tablo 7’de minimum destek 0,008, minimum güven ise 0,2 alınarak yapılan analiz sonuçları yer almaktadır. Tablo 7 incelendiğinde, Tablo 5 ve Tablo 6’da yer alan kurallardan farklı bir ilişkinin ortaya çıkmadığı görülmektedir. Bu sebeple Tablo 5 ve Tablo 6 üzerinden elde edilen sonuçlara göre web sitesinde yapılacak düzenlemelerin kullanıcı ihtiyaçlarını karşılama noktasında yeterli olacağı belirtilebilir.

Tablo 6. Minimum destek kriteri 0,007 olarak yapılan analizlerdeki oluşan kurallar  
(Created rules by the analysis made as the minimum support criterion is 0,007)



Şekil 5. Aramalar arasındaki ilişkiler  
(The relationship between searches)

## 5. TARTIŞMA (DISCUSSION)

Minimum Güven: 0,2				
#	Aranan Kelime/Kelime Grupları	Yönelim	Aranan Kelime/Kelime Grupları	Tekrarlanma Miktarı
1	Yatay Geçiş Başvuruları	→	Yatay Geçiş	49
2	Akademik Takvim	→	Takvim	49
3	Takvim	→	Akademik Takvim	49
4	Kayıt Yenileme	→	Harç	48
5	Kayıt Yenileme	→	Ders Kayıtları	48
6	Kontenjanlar	→	Yatay Geçiş	40

Minimum Güven: 0,4				
#	Aranan Kelime/Kelime Grupları	Yönelim	Aranan Kelime/Kelime Grupları	Tekrarlanma Miktarı
7	Yatay Geçiş Başvuruları	→	Yatay Geçiş	49
8	Takvim	→	Akademik Takvim	49

Tablo 7. Minimum destek kriteri 0,008 olarak yapılan analizlerdeki oluşan kurallar  
(Created rules by the analysis made as the minimum support criterion is 0,008)

Minimum Güven: 0,2				
#	Aranan Kelime/Kelime Grupları	Yönelim	Aranan Kelime/Kelime Grupları	Tekrarlanma Miktarı
1	Yatay Geçiş Başvuruları	→	Yatay Geçiş	49
2	Akademik Takvim	→	Takvim	49
3	Takvim	→	Akademik Takvim	49
4	Kayıt Yenileme	→	Harç	48
5	Kayıt Yenileme	→	Ders Kayıtları	48
6	Kontenjanlar	→	Yatay Geçiş	40

Elde edilen bulgular ışığında yapılan analiz değerlendirmeleri görsel biçimde ifade edilecek olursa Şekil 5’teki gibi bir ilişki ortaya çıkmaktadır.

Bu çalışmada, web sitesi ziyaretçilerinin site-içi aramalarında doğrudan kullandıkları kelime/kelime

gruplarının apriori algoritmasıyla analiz edilerek ihtiyaçlarının tespit edilebilmesi ve web sitesinin bu doğrultuda geliştirilebilmesi için bir örnek olay incelemesi sunulmuştur. Bu incelemede, analizler Kırklareli Üniversitesi'nin 2017 Ağustos ayına ait web günlük dosyalarından elde edilen veri seti üzerinde gerçekleştirilmiştir.

Sepet analizlerinde ürün tanımlarının/kodlarının belirli olmasından dolayı ürün satış ilişkilerinin ortaya çıkarılmasında apriori algoritması oldukça kullanışlıdır. Bu çalışmada ise, ziyaretçiler tarafından yapılan aramalardaki kelime/kelime gruplarının çeşitliliği bir dezavantaj olarak ortaya çıkmıştır; ancak OpenRefine uygulaması sayesinde site-içi aramalarda yazılan kelimeler/kelime grupları düzenlenerek bu dezavantajın üstesinden gelinmeye çalışılmıştır. Yine ziyaretçi aramalarındaki çeşitlilik, apriori algoritmasıyla kurallar oluşturulurken denenmiş olan destek ve güven kriterlerinin düşük değerlere sahip olmasının sebebi oluşturmuştur. Benzer kullanıcı aramalarının daha farklı yöntemlerle belirli kalıplar altında toplanması sayesinde destek ve güven kriterlerinin artırılması sağlanabilir.

Çalışmada ortaya çıkan bulgular ışığında, üniversite web sitesinde “Kayıt Yenileme” sürecinin açık bir şekilde ifade edilmesi ve “Harç” ile “Ders Kayıt” işlemlerinin nasıl gerçekleştirildiğine yönelik bilgilerin de bu sürecin açıklamalarında yer alması gerekliliği belirlenmiştir. Bununla birlikte, farklı üniversitelerden gelecek olan öğrencilere yönelik “Yatay Geçiş” ve “Kontenjan” bilgilerinin web sitesinde kullanıcının daha rahat ulaşabileceği belirgin bir noktaya taşınması gereği tespit edilmiştir. Ayrıca “Ders Programı” bilgilerinin de destekleyici olarak her iki ihtiyacın açıklamalarında bulunmasının uygun olacağı belirlenmiştir. “Akademik Takvim”in, yapılan diğer aramalarla ilişkisi bulunmamasına rağmen (“Takvim” aramasının da aynı amaçla yapıldığı düşünülmektedir), ilgili ay içinde en sık tekrar eden 10 arama içinde yer alması sebebiyle, “Akademik Takvim” ziyaretçilerin ihtiyaç duyduğu bir içerik olarak nitelendirilebilir. Tablo 5, Tablo 6 ve Tablo 7’den elde edilen kurallar dikkate alınarak birlikte aratılan kelimeler birbirine yakın biçimde web sitesine yerleştirilebilir. Tüm bu ihtiyaçların, uygun bir şablonla sunulacak olan farklı web sayfaları, grafikler, vb. araçlarla giderilebileceği söylenebilir.

Ayrıca, ziyaretçilerin arama yaptıkları sayfa üzerinde yazdıkları kelimeye yönelik otomatik tamamlama özelliğinin eklenmesi sayesinde, kullanıcılar tarafından yapılması muhtemel yazım yanlışlarının önüne geçmek mümkündür. Bu sayede, analiz gerçekleştirme aşamasında, aramalar arasındaki ilişki tespiti gerçekleştirilirken test edilebilecek destek ve güven değerlerinin artırılması sağlanabilir. Diğer bir ifadeyle; yazım hatalarının aza indirgenmesiyle ziyaretçilerin aramaları belirli bir kalıp altında toplanabilir. Otomatik tamamlama özelliğinin, ziyaretçilerin aradıkları içeriği bulmaya çalıştıkları noktada daha az emek harcamalarına ve web sayfasının

daha verimli çalışmasına katkı sağlayacağı düşünülmektedir. Dahası; birlikte aranan kelimeler, otomatik tamamlama özelliğine entegre edilebilir. Böylelikle Tablo 7’deki “Kayıt Yenileme” → “Harç” kuralından yola çıkılarak, arama kutusuna “Kayıt Yenileme” yazan kullanıcıya hem “Kayıt Yenileme” hem de “Harç” kelime/kelime grupları bir arada döndürülebilir.

Analizlerin sadece tek bir ay üzerinden gerçekleştirilmesi çalışmanın sınırlılıklarından biri olarak görülmüştür. İleriye yönelik gerçekleştirilecek çalışmalarda zaman aralığı geniş tutularak destek ve güven değeri daha yüksek kuralların elde edilmesi mümkündür. Analiz sonuçlarına göre web sitesinde gerçekleştirilecek olan değişikliklerin ne kadar etkili olduğu ölçülebilir. Yine, ileriki çalışmalarda apriori algoritması dışındaki birliktelik kuralı algoritmalarından da faydalanarak analizler gerçekleştirilebilir.

## 6. SONUÇ (CONCLUSION)

Zhang ve diğerleri [25] tarafından da benzer şekilde belirtildiği gibi web günlük dosyasının birliktelik kurallarıyla analiz edilmesi, web sitelerinin kişiselleştirilmesine ve daha iyi hizmet vermesine katkı sağlayacak tespitler yapma imkânı sunmaktadır. Bu çalışmada da örnek durum incelemesi üzerinden ele alınan yöntem; hem Kırklareli Üniversitesi'nin hem de diğer üniversitelerin web sitesi yöneticilerine, web sitesini içerik ve tasarım bakımından geliştirmelerinde fayda sağlayacak bulgular ortaya koyduğu tespit edilmiştir.

Kırklareli Üniversitesi'nin güz dönemi yeni kayıt/kayıt yenileme işlemlerinin gündemde olduğu Ağustos ayına ait web günlük dosyalarının analiziyle elde edilen bulgular sayesinde, web sitesinde yapılacak iyileştirmeler, bahar dönemi başlangıcında da web sitesinde ihtiyaç duyulacağı düşünülen içeriklere daha hızlı erişim olanağı tanıyacaktır. Böylelikle, web sitesinin daha verimli kullanılmasına yönelik olumlu katkıların sağlanacağı düşünülmektedir. “Akademik Takvim”, “Kayıt Yenileme”, “Başvuru” gibi üniversitelere ait benzer başlıklı içeriklerin mevcudiyeti gözönünde bulundurulursa, Kırklareli Üniversitesi'nden yola çıkarak oluşturulan önerilerin diğer üniversite web sayfaları için de geçerli olabileceğini söylemek mümkündür. Bu noktadan yola çıkılarak, bir üniversite web sitesinin iyileştirilmesine yönelik getirilen öneriler, “Dönemsel” ve “Sürekli” güncelleme önerileri olmak üzere iki ana başlıkta toplanmıştır. Dönemsel güncellemeler, her ne kadar bu çalışmada yalnızca güz dönemi temelinde oluşturulmuş olsa da, hem güz hem de bahar dönemlerinde farklı zaman aralıklarında gerçekleştirilen analizler sonucunda ihtiyaçların belirlenmesiyle web sitesinde iyileştirme ya da geliştirmeler yapılabilir. Talabeigi, Forsati ve Meybodi [47] tarafından da vurgulandığı gibi, kullanıcı eğilimlerine yönelik önerilerin yeterli olup olmadığı dönemlik analizlerle sağlanabilir. Dolayısıyla güncellemelere yönelik analizlerin belirli zaman aralıklarında yapılması, web sitelerinin daha iyi hizmet vermesi açısından

önemlidir. Web sitesinde gerçekleştirilecek olan sürekli güncellemeler ise, web sitesinde her zaman aktif olması gerektiği düşünülen güncelleme önerilerine karşılık gelmektedir.

#### Dönemsel Öneriler:

- Güz ve bahar dönemlerinin başlangıcında “Kayıt Yenileme”, “Harç” ve “Ders Kayıt” (A aramaları) ile ilgili içeriklerin tek bir sayfada toplanarak web sitesinde açılır bir pencereyle sunulması sağlanabilir. Böylelikle ziyaretçilerin sıklıkla talep ettiği bu içerikler, ziyaretçiler web sitesine ilk giriş yaptığında hemen karşlarına getirilecektir. “Yatay Geçiş” ve “Kontenjan” (B aramaları) başlıkları için de benzer yöntem uygulanabilir.
- Yukarıda verilen öneriye alternatif olarak, A aramaları için “Kayıtlar”, B aramaları içinse “Yatay Geçiş” başlığıyla ayrı ayrı birer küçük resim hazırlanabilir. Bu küçük resimlere, ilgili içeriğe yönlendiren birer bağlantı eklenerek ana sayfada ziyaretçilerin dikkatini çeken uygun bir noktaya (örneğin; ana sayfada yer alan üst menünün hemen altına, sayfanın sol en üst bölümüne) yerleştirilebilir.

#### Sürekli Öneriler:

- “Akademik Takvim” araması için ilgili içeriğe yönlenen bir bağlantıyı içeren küçük bir resmin, yine web sayfası ziyaretçilerinin dikkatini çekecek uygun bir noktaya (örneğin; ana sayfada yer alan üst menünün hemen altına, sayfanın sol en üst bölümüne) yerleştirilmesi sağlanabilir.
- “Ders Programı” aramasıyla ilgili olarak üniversitedeki tüm bölüm ve anabilim dallarındaki ders programlarını içeren genel bir web sayfası hazırlanabilir. Daha sonra üniversite web sitesindeki menüde yer alan “Akademik” ve “Öğrenci” bağlantılarına alt bağlantı eklenerek hazırlanan web sayfasına yönlendirme yapılabilir. Böylelikle, hem farklı bölümlerde çalışan personele hem de öğrencilere tek bir yerden ilgili döneme ait ders programına ulaşma imkânı sunulabilir.
- Yapılan aramaların standartlaştırılması amacıyla web sitesinde yer alan site-içi arama aracına otomatik tamamlama özelliği eklenebilir. Bu özelliğin çalışması için ziyaretçi aramalarının bir veri tabanı oluşturularak depolanması gerekmektedir. Bu sayede, ziyaretçi herhangi bir içeriği aramaya çalıştığında ilk olarak veri tabanında benzer aramanın yapılma durumu kontrol edilecek; eğer yapıldıysa kullanıcıya ilgili sonuçlar getirilecek, aksi durumda hiçbir işlem gerçekleştirilmeyecektir. Veri tabanında mevcut olmayan aramalar gerçekleştirilince, arama havuzu genişletilebilecek, zamanla kullanıcıların daha

hızlı bir şekilde site-içi arama işlemi gerçekleştirmesi sağlanacaktır.

Çalışmanın odağında bir üniversite web sitesi olduğu için, yukarıda belirtilen önerilerin üniversiteler dışında hizmet vermekte olan diğer kurumların web sitelerine uyarlanması mümkün olmayacaktır. Ancak; çalışma yönteminde kullanılan adımların, farklı web sitelerinde ya da web sitesi dışında farklı uygulamalarda izlenmesi sonucunda, web sitelerinin/uygulamaların tasarım, içerik ve site-içi arama yönünden daha iyi hizmet vereceği düşünülmektedir.

#### 7. TEŞEKKÜRLER (ACKNOWLEDGEMENTS)

Yazarlar, çalışmada kullanılan web günlük dosyalarının kullanılmasını için gerekli iznin sağlanmasına yardımcı olan Gökhan Doğan’a ve çalışma bulgularının yorumlanmasına yaptığı değerli katkılardan ötürü Dr. Zeki Özen’e teşekkür eder.

#### KAYNAKLAR (REFERENCES)

- [1] Internet: InternetLiveStats.com, Total number of Websites - Internet Live Stats, <http://www.internetlivestats.com/total-number-of-websites/#trend>, 19.01.2018.
- [2] Y. İnal, N. Ö. Çınar, ve K. Çağlıtay, “Kamu İnternet Sitelerinde Yer Alan Arama Alanlarının Kullanılabilirliği ve Buna Yönelik Kullanıcı Davranışlarının Analizi”, *Bilişim Teknolojileri Dergisi*, 9(1), 41-54, 2016.
- [3] I. Graham, **A Pattern Language for Web Usability**, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2002.
- [4] B. Özkan ve Y. Özkan, **R ile Programlama**, 1, Papatya Yayıncılık Eğitim, İstanbul, Türkiye, 2017.
- [5] İ. E. Emre ve Ç. Selçukcan Erol, “Veri Analizinde İstatistik mi Veri Madenciliği mi?”, *Bilişim Teknolojileri Dergisi*, 10(2), 161-167, 2017.
- [6] A. Chauhan ve S. Tarar, “Prediction of User Browsing Behavior Using Web Log Data”, *International Journal of Scientific Research in Science, Engineering and Technology*, 2(1), 419-422, 2016.
- [7] I. Çınar ve H. Ş. Bilge, “Web Madenciliği Yöntemleri ile Web Loglarının İstatistiksel Analizi ve Saldırı Tespiti”, *Bilişim Teknolojileri Dergisi*, 9(2), 125-135, 2016.
- [8] J. Grace, V. Maheswari, ve N. Dhinakaran, “Analysis of Web Logs And Web User In Web Mining”, *International Journal of Network Security & Its Applications*, 3(1), 99-110, 2011.
- [9] A. Guerbas, O. Addam, O. Zaarour, M. Nagi, A. Elhadj, M. Ridley ve R. Alhadj, “Effective web log mining and online navigational pattern prediction”, *Knowledge-Based Systems*, 49(1), 50-62, 2013.
- [10] C. M. Barnum, “3 - Big U and little u usability”, **Usability Testing Essentials**, Morgan Kaufmann, Boston, ABD, 53-81, 2011.
- [11] J. Nielsen, **Usability Engineering**, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
- [12] M. Gezer, Ç. Erol, ve S. Gülseçen, “Bir Web Sayfasının Web Madenciliği İle Analizi”, **Akademik Bilişim**, Dumlupınar Üniversitesi, Kütahya, Türkiye, 1-10, 31 Ocak-2 Şubat, 2007.
- [13] C. Romero, S. Ventura, A. Zafra, ve P. de Bra, “Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems”, *Computers & Education*, 53(3), 828-840, 2009.

- [14] J.-H. Lee ve W.-K. Shiu, "An adaptive website system to improve efficiency with web mining techniques", *Advanced Engineering Informatics*, 18(3), 129–142, 2004.
- [15] E. Hochstain, "A Mining Approach to Evaluate Geoportals Usability", **2015 International Workshop on Data Mining with Industrial Applications (DMIA)**, Assuncion, Paraguay, 1–7, 14-16 Eylül 2015.
- [16] J. Srivastava, R. Cooley, M. Deshpande, ve P.-N. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data", *SIGKDD Explorations*, 1(2), 12–23, 2000.
- [17] R. Das ve İ. Türkoğlu, "Extraction of Interesting Patterns through Association Rule Mining For Improvement of Website Usability", *Istanbul University - Journal of Electrical & Electronics Engineering*, 9(2), 1037–1046, 2009.
- [18] N. K. Tyagi, A. K. Solanki, ve M. Wadhwa, "Analysis of Server Log by Web Usage Mining for Website Improvement", *International Journal of Computer Science Issues*, 7(4), 17-21, 2010.
- [19] K. R. Suneetha ve R. Krishnamoorthi, "Classification Of Web Log Data To Identify Interested Users Using Decision Trees", **The International Conference on Computing, Communications and Information Technology Applications**, 21-23 Ocak 2010.
- [20] C. J. Carmona, S. Ramirez-Gallego, F. Torres, E. Bernal, J. Del. ve S. García, "Web usage mining to improve the design of an e-commerce website: OrOliveSur.com", *Expert Systems with Applications*, 39(12), 11243–11249, 2012.
- [21] A. K. Santra ve S. Jayasudha, "Classification of Web Log Data to Identify Interested Users Using Naïve Bayesian Classification", *International Journal of Computer Science Issues*, 9(1), 381–387, 2012.
- [22] V. Sujatha ve Punithavalli, "Improved user Navigation Pattern Prediction Technique from Web Log Data", *Procedia Engineering*, 30, 92–99, 2012.
- [23] P. Lopes and B. Roy, "Dynamic Recommendation System Using Web Usage Mining for E-commerce Users", *Procedia Computer Science*, 45, 60–69, 2015.
- [24] M. Y. Shih ve S.-S. Huang, "Characterizing Web users based on their required criteria", **11th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QSHINE)**, Taipei, Tayvan, 422–426, 19-20 Ağustos 2015.
- [25] H. Zhang, W. Song, L. Liu, ve H. Wang, "The application of matrix Apriori algorithm in web log mining", **2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)**, Beijing, Çin, 264–268, 10-12 Mart 2017.
- [26] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining", *Journal of Data Warehousing*, 5, 13–22, 2000.
- [27] M. E. Balaban ve E. Kartal, **Veri Madenciliği ve Makine Öğrenmesi Temel Algoritmaları ve R Dili ile Uygulamaları**, 1, Çağlayan Kitabevi, İstanbul, Türkiye, 2015.
- [28] Z. Özen, E. Kartal ve İ. E. Emre, "Analysis of a Learning Management System by Using Google Analytics: A Case Study From Turkey", *Technology Management in Organizational and Societal Contexts*, IGI Global, ABD, 198–220, 2018.
- [29] Internet: Apache: The Apache HTTP Server Project, <http://httpd.apache.org/>, 23.01.2018.
- [30] D. B. Rathod, "Customizable Web Log Mining from Web Server Log", *International Journal of Engineering Development and Research*, 1(2), 96–100, 2014.
- [31] Internet: OpenRefine, <http://openrefine.org/>, 23.01.2018.
- [32] Internet: RStudio, Take control of your R code, <https://www.rstudio.com/products/rstudio/>, 23.01.2018.
- [33] Internet: r-project, R: The R Project for Statistical Computing, <https://www.r-project.org/>, 23.01.2018.
- [34] M. Hashler, S. Chelluboina, K. Hornik, ve C. Buchta, "The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Data Sets", *Journal of Machine Learning Research*, 12, 2021–2025, 2011.
- [35] Internet: R. Calaway, Microsoft, ve S. Weston, foreach: Provides Foreach Looping Construct for R, <https://cran.r-project.org/web/packages/foreach/foreach.pdf>, 12.12.2017.
- [36] Internet: R. Calaway, R. Analytics, ve S. Weston, doMC: Foreach Parallel Adaptor for "parallel", <https://cran.r-project.org/web/packages/doMC/doMC.pdf>, 12.12.2017.
- [37] H. Wickham, "The Split-Apply-Combine Strategy for Data Analysis", *Journal of Statistical Software*, 40(1), 1–29, 2011.
- [38] Internet: O. Keyes, J. Jacobs, D. Schmidt, M. Greenaway, B. Rudis, A. Pinto, M. Khezrzadeh, P. Meilstrup, A. M. Costello, J. Bezanson, P. Meilstrup ve X. Jiang, urltools: Vectorised Tools for URL Handling and Parsing, <https://cran.r-project.org/web/packages/urltools/urltools.pdf>, 20.01.2018.
- [39] R. Agrawal ve R. Srikant, "Fast Algorithms for Mining Association Rules", **Proceedings of the 20th International Conference on Very Large Data Bases**, Santiago, Şili, 487–499, 12-15 Eylül, 1994.
- [40] G. Karahan Adalı, **Veri Madenciliğinde Birliktelik Yöntemleri ve Müşteri İlişkileri Yönetimine İlişkin Bir Uygulama**, Doktora Tezi, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, 2017.
- [41] Y. Zhao, **Post-mining of Association Rules: Techniques for Effective Knowledge Extraction**, 1, PA: Information Science Reference, Hershey, ABD, 2009.
- [42] C. Zhang ve S. Zhang, **Association Rule Mining**, Heidelberg: Springer Berlin Heidelberg, Berlin, ABD, 2002.
- [43] P.-N. Tan, M. Steinbach ve V. Kumar, **Introduction to Data Mining**, 1, Pearson, Boston, ABD, 2005.
- [44] M. Mittal, J. Singh, A. Aggarwal, K. Kumari, ve M. Yadav, "Ordering Policy for Imperfect Quality Itemssets Using Cross Selling Effects", *International Journal of Modeling and Optimization*, 4(1), 25-30, 2014.
- [45] Internet: draw.io, Flowchart Maker and Online Diagram Software, <https://www.draw.io/>, 03.04.2018.
- [46] Internet: J. Nielsen, F-Shaped Pattern For Reading Web Content (original study), <https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content-discovered/>, 28.03.2018.
- [47] M. Talabeigi, R. Forsati ve M.R. Meybodi, "A Dynamic Web Recommender System Based on Cellular Learning Automata", **2010 International Conference on Computer Engineering and Technology**, Chengdu, Çin, 755-761, 16-18 Nisan 2010.