

Karar Ağaçları ile Bir Web Sitesinde Link Analizi ve Tespiti

Muhammed Aksu, Ersin Karaman

ÖZET

Web siteleri metinler, resimler, videolar gibi içerikler ile ziyaretçilerine bilgi vermeyi amaçlayan uygulamalardır. Web sitesi ana sayfasının sitenin bir özeti olması ve çok tıklanan, çok zaman geçirilen ya da hemen çıkma oranı düşük olan alt sayfaların linklerini barındırması beklenmektedir. Büyük web siteleri yüzlerce ya da binlerce alt sayfa içermektedir. Bu alt sayfaların hepsinin linklerinin ana sayfaya konulması, ana sayfanın karmaşık bir hal almasına neden olur. Bu çalışmada, bir web sitesinin Google Analitik verilerini kullanarak, veri madenciliği yöntemiyle ana sayfada olması gereken linklerin tespitini yapabilmek için bir karar ağacı oluşturulmuştur. Oluşturulan karar ağacı mevcut ya da yeni yapılacak bir web sitesinin ana sayfasında olması gereken alt sayfalarının linklerinin tespitinde kullanılabilir. Böylelikle ana sayfanın daha işlevsel olması, ziyaretçilerin girmek istedikleri alt sayfaları daha hızlı bulmaları sağlanabilir.

Anahtar Kelimeler: Veri Madenciliği, Google Analitik, WEKA, Sınıflandırma, C4.5

SUMMARY

Web sites are webbased applications that intend to inform their visitors through contents such as texts, visuals and videos. The home page of a web site must be a summary of the web site. Home page must host the links for frequently or longterm visited subpages with lower possibility to exit immediately. Big web sites have hundreds or thousands of subpages. If the links of all these subpages are given in the home page, the page can become messed. In this research, a decision tree was formed through data mining method using Google Analytic data in order to identify the links necessary to be given in the home page of a web site. The decision tree can be used for the identification of the links of the subpages necessary to be given in the home page of a current or future web site. This can support the functionality of the home page and help the visitors easily find the subpages they would like to enter into.

Keywords: Data Mining, Google Analytic, WEKA, Classification, C4.5

Information of Author(s):

Muhammed Aksu*
m.cagri.aksu@gmail.com

Ersin Karaman
karamanersin@gmail.com

Submit Date: 20.11.2017

Accept Date: 21.12.2017

Publish Date: 29.12.2017



* Contact Author

Address:

Telephone Number: 0466 215 10 00 / 1064

E-mail: m.cagri.aksu@gmail.com

1.GİRİŞ

Web siteleri metinler, resimler, videolar gibi içerikler ile ziyaretçilerine bilgi vermeyi amaçlayan uygulamalardır. İnternette kişisel web sitesi, kurumsal web sitesi, eticaret web siteleri gibi çeşitli web siteleri vardır. Kurumsal web siteleri bir kurum ya da kuruluş hakkında bilgi veren metinleri, kurum ya da kuruluşun haberlerini, duyurularını, etkinliklerini barındıran, genellikle o kurum ya da kuruluş ile alakalı kişilerin ziyaret ettiği web siteleridir.

Kurumsal web siteleri genellikle bir ana sayfa ve bu ana sayfaya bağlı birçok alt sayfalardan oluşmaktadır. Kurumsal bir web sitesinin ana sayfası, sitenin bir özeti şeklinde olmalıdır. Çok tıklanan, çok zaman geçirilen ya da hemen çıkma oranı düşük olan alt sayfaların linklerinin ana sayfada olması, web sitesinin işlevselliği açısından önem arz etmektedir. Büyük web sitelerinde yüzlerce, belki binlerce alt sayfa bulunmaktadır. Bu alt sayfaların her birine ana sayfada link vermek, ana sayfanın karmaşık bir hal almasına neden olmaktadır, bu da kullanıcıların aradıkları alt sayfayı bulmalarını zorlaştırmaktadır. Bu nedenle ana sayfada kullanıcıların çok ziyaret ettiği, çokça zaman geçirdiği, sayfadan hemen çıkmadığı alt sayfaların linklerinin bulunması gerekmektedir. Bu çalışmada Artvin Çoruh Üniversitesi web sitesinin Google Analitik verileri çeşitli karar ağacı algoritmalarıyla sınıflandırılmıştır. Sınıflandırma neticesinde en iyi sonucu veren algoritma ile bir karar ağacı oluşturulmuştur. Oluşturulan karar ağacı mevcut web sitesinin ana sayfasında bulunması gereken linklerin tespitinde kullanılmıştır.

Literatüre bakıldığında web sitelerinin fonksiyonelliğini artırmak için Google Analitik verilerinden ziyade web erişim günlükleri kullanılmıştır. Resul, Türkoğlu ve Poyraz bir web sitesinin erişim kayıtlarının analizini, web kullanım madenciliği ile gerçekleştirmişlerdir [1]. Daş, Türkoğlu çalışmalarında web kayıt dosyalarından ilginç örüntüleri keşfetmek konusu hakkındaki çalışmaları incelemişlerdir [2]. Tuğ, Şakiroğlu yaptıkları çalışmalarında genetik algoritma yöntemiyle web erişim kayıtlarından sıralı ilişkileri tespit etmişlerdir [3]. Carmona, RamírezGallego bir eticaret web sitesinin Google Analitik verileriyle çeşitli veri madenciliği işlemleri yapmışlardır [4]. Neticesinde web sitesinin kullanıcı memnuniyetini artıracak sonuçlar elde etmişlerdir. Parvatikar and Joshi web kayıt dosyalarını kullanarak çeşitli veri madenciliği teknikleriyle son kullanıcı ile ilgili bilgiler ortaya çıkarmışlardır [5].

2.YÖNTEM

Bu çalışmada Artvin Çoruh Üniversitesi web sitesinin 01.01.2015 - 01.01.2017 tarihleri arasındaki Google Analitik verileri kullanılmıştır. Bu veriler çeşitli sınıflandırma algoritmalarıyla sınıflandırılmış ve neticesinde en iyi sınıflandırma başarısına sahip olan algoritma ile bir karar ağacı oluşturulmuştur. Oluşturulan karar ağacı ile üniversitenin ana sayfasında olması gereken linklerin tespit edilmesi amaçlanmıştır. Çalışmada kullandığımız veriler içerisinde web sitesinde bulunan sayfaların, sayfa görüntüleme sayısı, tekil sayfa görüntüleme sayısı, sayfada kalma süresi, giriş sayıları, çıkış oranı, hemen çıkma oranı alanları bulunmaktadır. Tablo 1’de verilerin bir örneği gösterilmektedir. 21371 adet veri içerisinde duyuru linkleri, etkinlik linkleri, haber linkleri gibi yapılacak işlemlerle alakalı olmayan veriler temizlenmiş geriye 2536 adet veri kalmıştır.

Tablo 1. Google Analitik Veri Örneği

Sayfa Linki	Sayfa Gör. Sayısı	Tekil Sayfa Gör. Sayısı	Ort. Kal. Süre.	Giriş Sayısı	Çıkış Oranı	Hem. Çık. Oranı
/tr/menu/137/131/egitim-fakultesi	140585	84038	44	20720	13	22
/tr/menu/145/134/artvin-my	107951	70164	62	26663	24	36

/tr/menu/131/19/fakulteler	106666	73313	14	1699	5	26
/tr/menu/147/134/hopa-myo	80020	52797	57	30766	26	32
/tr/menu/19/16/akademik	79796	51726	11	987	4	32
/tr/menu/1018/968/idari	1431	912	113	10	13	30

Veriler temizlendikten sonra geliştirilen küçük bir web tabanlı program ile bu 2536 adet verinin hangilerinin linklerinin, mevcut ana sayfada buldukları tespit edilmiştir. Bu web tabanlı uygulama mevcut web sitesinin ana sayfasının kaynak koduna bakmakta, Tablo 1’de örneği gösterilen verilerin sayfa linki alanını kaynak kod ile karşılaştırmakta, kaynak kod içerisinde sayfanın linkini tespit ederse bu verinin ana sayfada bulunup bulunmama alanını “true” olarak belirlemekte, eğer kaynak kod içerisinde sayfanın linkini tespit edemezse “false” olarak belirlemektedir. Böylece tüm verilere bir alan daha eklenmiştir. Bu alan sayfanın ana sayfada linkinin bulunup bulunmadığını gösteren bir alandır. Eğer sayfanın ana sayfada linki bulunuyorsa “true”, bulunmuyorsa “false” değerini içermektedir.

Tablo 2. Veri Madenciliğine Hazır Hale Getirilmiş Verilerin Örneği

Sayfa Gör. Sayısı	Tekil Sayfa Gör. Sayısı	Ort. Kal. Süre.	Giriş Sayısı	Çıkış Oranı (%)	Hem. Çık. Oranı (%)	Ana Sayf. Linki Var Mı?
140585	84038	44	20720	13	22	True
107951	70164	62	26663	24	36	True
106666	73313	14	1699	5	26	True
80020	52797	57	30766	26	32	True
79796	51726	11	987	4	32	True
1431	912	113	10	13	30	False

Veriler, yukarıda belirtildiği gibi düzenlendikten sonra Tablo 2’de görüldüğü gibi veri madenciliği yapılacak hale getirilmiştir. Daha sonra veriler, veri madenciliği uygulamalarında sıklıkla kullanılan Weka programı vasıtasıyla, karar ağacı algoritmalarından C4.5, Random Forest, REP Tree, Logistic Model Tree ve CART algoritmaları kullanılarak sınıflandırılmıştır. Sınıflandırma başarısı en yüksek olan C4.5 algoritması ile karar ağacı oluşturulmuştur. Oluşturulan bu karar ağacıyla ana sayfada bulunması gereken linklerin tespit edilmesi amaçlanmıştır. Bu karar ağacı mevcut web sitesi için kullanılacağı gibi yeniden yapılması planlanan site için de kullanılabilir. Böylelikle web sayfasının daha fonksiyonel ve kullanışlı olması sağlanabilir.

Bu kısımdan sonra, yöntemin alt başlıklarında, çalışmanın yöntemi olan veri madenciliği, karar ağaçları, Google Analitik, çalışmada kullanılan karar ağacı algoritmaları ve Weka programı hakkında bilgiler verilmektedir.

2.1. Veri Madenciliği

Bilişim teknolojilerindeki gelişmeler neticesinde, artık her veri sayısal ortama kaydedilmektedir ve veri tabanlarında milyonlarca veri biriktirmektedir. Biriken bu veriler saklandığı yerde öylece beklerken bir anlam ifade etmemektedirler. Bu verilerin anlam ifade eden yararlı bilgilere dönüştürülmesi gerekmektedir. Verilerin anlam ifade eden bilgilere dönüştürülmesi veri madenciliği vasıtasıyla olmaktadır [6]. Aşağıda veri madenciliğinin ne olduğunu gösteren farklı tanımlar verilmiştir.

- Veri madenciliği daha önceden bilinmeyen, geçerli ve uygulanabilir bilgilerin geniş veri tabanlarından elde edilmesi ve bu bilgilerin işletme kararları alınırken kullanılmasıdır [7].
- Veri madenciliği büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanarak aranmasıdır [8].

Veri madenciliğinde birçok yöntem ve algoritma bulunmaktadır, ancak genel olarak veri madenciliği sınıflandırma, kümeleme ve birliktelik kuralları olarak üç ana başlığa ayrılmaktadır. Sınıflandırma, sınıfı tanımlanmamış verileri kullanarak sınıfı belli olmayan verilerin sınıfını tahmin etmek için kullanılan bir yöntemdir. Sınıflandırmanın ilk aşamasında veriler herhangi bir sınıflandırma algoritmasına tabi tutulmakta ve kurallar oluşturulmaktadır. Sınıflandırmanın ikinci aşamasında ise ilk aşamada oluşturulan kurallar, sınıfı belli olmayan veriler üzerine uygulanmakta ve verilerin sınıfı tahmin edilmektedir [9]. Kümelemede, verilerin benzerliklerine göre kümelenebilecek işlemleri bulunmaktadır [6]. Birliktelik kurallarında ise veriler içerisinde bulunan birliktelik ilişkilerinin çıkarılması işlemi yapılmaktadır. Veriler içerisindeki tüm kombinasyonlara bakılmakta, birbiri ile ilişkili olan veriler keşfedilmeye çalışılmaktadır [10].

2.2.Karar Ağaçları

Karar ağaçları, veri madenciliğinde sınıflandırma tekniklerinden birisidir. Sınıflandırma problemlerinde çokça kullanılmaktadır. Karar ağaçlarında önce bir karar ağacı oluşturulmakta, daha sonra karar ağacından üretilen kurallar ile veri tabanında bulunan kayıtlar sınıflandırılmaktadır [7].

Karar ağaçlarında ID3, C4.5, Sliq, Sprint, CART, REP Tree, Random Forest, Logistic Model Tree gibi birçok algoritma kullanılabilmektedir. Veriler önce bu karar ağacı algoritmalarından birine gönderilir. Algoritma bu verileri işler ve bir karar ağacı oluşturur. Oluşturulan bu karar ağacı sınıfı bilinmeyen veriler üzerine uygulanarak bu verilerin sınıflarının tespit edilmesi sağlanır.

2.3.C4.5 Algoritması

C4.5 algoritması 1993 yılında Quinlan tarafından ortaya çıkarılmış bir karar ağacı algoritmasıdır. Bu algoritma ile sayısal veriler karar ağaçlarında kullanılabilir hale gelmiştir. C4.5 algoritması yine Quinlan tarafından geliştirilen ID3 algoritmasının daha gelişmiş halidir. ID3 algoritmasında kayıp veriler hesaba katılmazken C4.5 algoritmasında kayıp verilerin değerleri tahmin edilerek kullanılmaktadır [6].

C4.5 algoritması sınıflandırma işlemini verilerin sıklıklarına göre yapmaktadır. Veri kümesindeki düzensizliği gösteren entropi hesabına bağlı olarak çalışmaktadır. C4.5 algoritmasının çalışma şekli aşağıdaki gibi açıklanabilir.

- $\sum_{i=1}^n p_i \log_2 p_i$ formülüne göre veri kümesindeki tüm sütunların entropisi hesaplanır [11].
- Veri kümesindeki her bir sütunun entropisi, sınıfın entropisine bölünerek her bir sütunun kazancı hesaplanır.
- Kazancı en yüksek olan kök düğüm olarak atanır.

Kök düğümün ait olduğu veriler haricinde kalan veriler için aynı işlemler tekrarlanır. Tüm veriler bitene kadar aynı işlemler devam eder [12].

2.4.REP Tree Algoritması

Bu algoritma ilk olarak Quinlan tarafından önerilmiştir [13]. REP Tree algoritmasında, regresyon ağacı mantığıyla, farklı yinelemelerde birden fazla ağaç oluşturulmakta ve daha sonra oluşturulan bu karar ağaçlarından en iyisini seçilmektedir. Algoritma varyansdan kaynaklanan hatayı en aza indirme ilkesine ve entropi ile bilgi kazanımı ilkesine dayanmaktadır [14-16].

2.5.Random Forest Algoritması

Bu algoritma Breiman tarafından ileri sürülmüştür. Özellikle büyük miktarda verilerin sınıflandırılmasında iyi performans göstermektedir. Amacı tek bir karar ağacı üretmek yerine her biri farklı eğitim kümelerinde eğitilmiş

olan çok sayıda karar ağacını birleştirmektedir. Farklı eğitim kümeleri oluştururken ön yükleme ve rastgele özellik seçimi kullanılmaktadır [17, 18].

2.6.CART Algoritması

1984 yılında Breiman tarafından oluşturulmuş bir algoritmadır. CART algoritmasında karar ağaçları, her bir karar düğümünün çeşitli ayrılma kriterleri ile iki farklı dala ayrılmasıyla oluşturulmaktadır [6, 19].

2.7.Logistic Model Tree Algoritması

Logistik Model Tree, kısaca LMT, indiksiyon ağaçlarının ve lojistik regresyonun birleştirilmesiyle oluşmuş bir algoritmadır. Bu algortmada, ağacın her düğümünde bir lojistik regresyon modeli üretmek için LogitBoost algoritması kullanılmakta, daha sonra düğümler C4.5 kriterleri kullanılarak bölünmektedir [20, 21].

2.8.Google Analitik

Google Analitik, dünyanın en büyük arama motoru olan Google'ın web sitesi sahipleri için sunduğu web sitesi analiz aracıdır. Google analitik ile web sitesi ziyaretçilerinin, web sitesine nasıl ulaştıkları ve sitede ne gibi işlemler yaptıkları kolayca analiz edilebilmektedir. Google analitik ile elde edilen bilgiler ile site tasarımı geliştirilebilmekte, girilen içeriklerin ilgi çekip çekmediği anlaşılabilenmekte ve web sitesi daha işlevsel hale getirilebilmektedir [22].

Google analitik ile belirttiğiniz tarihler arasında web sitenize gelen ziyaretçilerin sayısı görülebileceği gibi anlık ziyaretçi sayıları da görülebilmektedir. Ayrıca Google Analitik ile web sitenizde her bir sayfanın detaylı istatistiksel bilgisi alınabilmektedir. Google Analitiğin sunduğu istatistiksel veriler içerisinde sayfa görüntüleme sayısı, tekil sayfa görüntüleme sayısı, sayfada geçirilen süre, giriş sayısı, çıkış oranı, hemen çıkma oranı, hangi tarayıcı ile giriş yapılmış, hangi şehirden giriş yapılmış, hangi arama kelimesi kullanılarak giriş yapılmış gibi birçok alan bulunmaktadır. Bu çalışmada, Google Analitiğin sunduğu bu alanlardan, bir sayfanın ana sayfada linkinin olmasını gerektirecek alanlar olan sayfa görüntüleme sayısı, tekil sayfa görüntüleme sayısı, sayfada geçirilen süre, giriş sayısı, çıkış oranı ve hemen çıkma oranı alanları kullanılmıştır. Sayfa görüntüleme sayısı, o sayfaya kaç kere girildiğini göstermektedir. Tekil sayfa görüntüleme sayısı, o sayfaya kullanıcı bazlı olarak girişi göstermektedir. Giriş sayısı, kullanıcının web sitesine ilk giriş yaptığı sayıdır. Örneğin kullanıcı A sayfasına giriş yapıp aynı site içerisinde B sayfasına geçiş yaparsa A sayfasının giriş sayısı değeri 1 artarken B sayfasının giriş sayısı değişmez. Hemen çıkma oranı, sitedeki tek sayfalık oturumlara denirken, çıkma oranında bir oturumda birden fazla sayfa gezerek çıkış işlemi bulunmaktadır.

2.9.Weka

Weka, veri madenciliği işlemleri için kullanılan, makine öğrenme algoritmalarını barındıran, Yeni Zelanda'daki Waikato Üniversitesindeki makine öğrenmesi grubu tarafından geliştirilmiş açık kaynak kodlu bir programdır. Weka ile veri madenciliği algoritmaları doğrudan bir veri kümesine uygulanabilmekte ya da geliştirilen java kodundan çağrılabilir. Weka programında, veri ön işleme, sınıflandırma, kümeleme, ilişkilendirme kuralları ve görselleştirme araçları bulunmaktadır. Ayrıca yeni makine öğrenme planları geliştirmek için de uygun bir programdır [23].

3.Bulgular

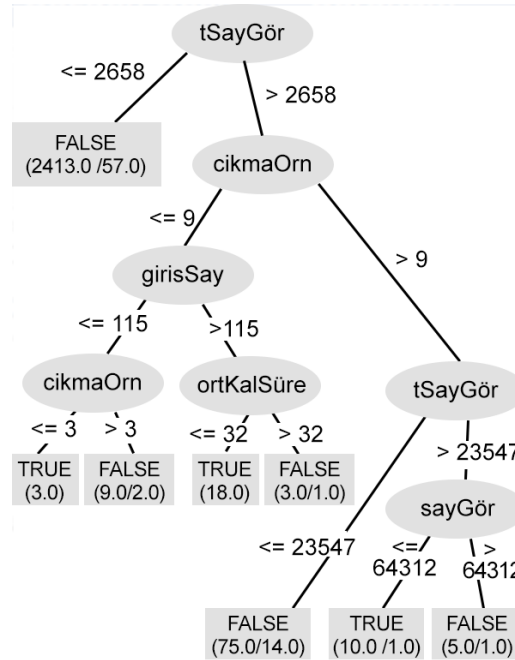
Karar ağaçları yönteminde sınıflandırılması istenilen veriler, sınıflandırma algoritmasına gönderilmekte ve sınıflandırma sonucunda bir karar ağacı oluşturulmaktadır. Oluşturulan karar ağacına göre sınıfı belli olmayan verilerin sınıflandırma işlemi gerçekleştirilmektedir.

Tablo 3. Algoritmaların Doğru Sınıflandırma Oranları

Algoritma	Doğru Sınıflandırma Oranı
C4.5	96.3722 %
Random	96.2934 %
REP Tree	96.2145 %
CART	96.1356 %
LMT	95.9779 %

Bu çalışmada yöntem kısmında bahsedilen veriler weka programı vasıtasıyla sınıflandırma algoritmalarından C4.5, Random Forest, REP Tree, Logistic Model Tree ve CART algoritmalarıyla sınıflandırılmıştır. Sınıflandırma işlemi sonucunda Tablo 3’de de görüldüğü gibi en yüksek sınıflandırma oranına C4.5 algoritması ulaşmıştır. En başarılı sınıflandırma işlemi C4.5 algoritması yaptığı için C4.5 algoritmasının karar ağacı oluşturulmuştur. Şekil 1’de bu karar ağacı gösterilmektedir. Bu karar ağacına bakılarak aşağıdaki karar kuralları oluşturulabilmektedir.

1. Tekil Sayfa Görüntüleme Sayısı ≤ 2658 ise Ana Sayfada Göster = False
2. Tekil Sayfa Görüntüleme > 2658 ve Çıkma Oranı ≤ 9 ve Giriş Sayısı ≤ 115 ve Çıkma Oranı ≤ 3 ise Ana Sayfada Göster = True
3. Tekil Sayfa Görüntüleme > 2658 ve Çıkma Oranı ≤ 9 ve Giriş Sayısı ≤ 115 ve Çıkma Oranı > 3 ise Ana Sayfada Göster = False
4. Tekil Sayfa Görüntüleme > 2658 ve Çıkma Oranı ≤ 9 ve Giriş Sayısı > 115 ve Ortalama Sayfada Kalma Süresi ≤ 32 ise Ana Sayfada Göster = True
5. Tekil Sayfa Görüntüleme > 2658 ve Çıkma Oranı ≤ 9 ve Giriş Sayısı > 115 ve Ortalama Sayfada Kalma Süresi > 32 ise Ana Sayfada Göster = False
6. Tekil Sayfa Görüntüleme > 2658 ve Çıkma Oranı > 9 ve Tekil Sayfa Görüntüleme Sayısı ≤ 23547 ise Ana Sayfada Göster = False
7. Tekil Sayfa Görüntüleme > 2658 ve Çıkma Oranı > 9 ve Tekil Sayfa Görüntüleme Sayısı > 23547 ve Sayfa Görüntüleme Sayısı ≤ 64312 ise Ana Sayfada Göster = True
8. Tekil Sayfa Görüntüleme > 2658 ve Çıkma Oranı > 9 ve Tekil Sayfa Görüntüleme Sayısı > 23547 ve Sayfa Görüntüleme Sayısı > 64312 ise Ana Sayfada Göster = False



Şekil 1. Çalışma Sonucunda Oluşturulan Karar Ağacı

4. Tartışma ve Sonuç

Weka programı birçok veri madenciliği yöntemini desteklemekte ve her bir yönteme ait pek çok algoritma ile analiz yapabilmektedir. Weka ile veri madenciliği algoritmaları doğrudan veri kümesine uygulanarak veri madenciliği işlemi gerçekleştirilebilmektedir. Bu çalışmada Artvin Çoruh Üniversitesinin web sitesine ait 2536 adet Google Analitik verisi weka programı kullanarak çeşitli sınıflandırma algoritmaları kullanılarak sınıflandırılmış ve en iyi sınıflandırma işlemi yapan algoritma kullanılarak 8 adet kural üreten bir karar ağacı oluşturulmuştur. Bu kurallar kullanılarak bir alt sayfanın linkinin ana sayfada bulunmasının gerekli olup olmadığı anlaşılabilir. Böylece web sitesinin ana sayfası, daha çok görüntüleme alan, sayfada kalma süresi yüksek olan, çıkma oranı düşük olan sayfaları barındırır ve web sitesini kullanan kullanıcılar için aradıkları bilgiye erişmeyi kolaylaştırır ayrıca web sitesinin daha fonksiyonel olması sağlanır.

Uygulama sonucuna bakılacak olursa, verilerin Google Analitikten alınıp weka programıyla veri madenciliği işlemine tabi tutulup, sonucunda karar ağacı üretilip ve bu karar ağacından karar kuralları üretmek oldukça zahmetli ve iş yükü gerektiren işlemlerden oluşmaktadır. Bu işlemlerin hepsi tek bir web tabanlı uygulama ile rahatlıkla gerçekleştirilebilir. Geliştirilecek web tabanlı uygulama Google Analitik servislerini kullanarak web sitesine ait verileri Google Analitikten çekebilir, çektiği verileri veri madenciliğine hazır hale getirdikten sonra bu uygulamada kullandığımız C4.5 algoritmasına göndererek karar ağacını oluşturabilir. Karar ağacını oluşturduktan sonra verilen linkin ana sayfada olup olmadığına karar verebilir. Böylece karar ağacı üretme ve link tespiti işlemi çok daha hızlı bir şekilde gerçekleştirilebilir.

Bu çalışmada bir web sitesinin ana sayfasında bulunması gereken öğeler veri madenciliği yöntemi ile belirlenmiştir. Ancak bu verilerin insan bilgisayar etkileşimi açısından da incelenmesi sonuçların genellenebilirliği açısından önem taşımaktadır. Ayrıca bu çalışma, ana sayfa linklerinin belirlenmesinden sonra kullanıcı temelli bir değerlendirme yaparak kullanıcıların memnuniyetinde bir değişiklik olup olmadığını araştırılması sonraki çalışmalara yön verme noktasında önem arz etmektedir.

Kaynaklar

- [1] Resul, D., İ. Türkoğlu, ve M. Poyraz, Bir Web Sitesine Ait Kullanıcı Erişim Kayıtlarının Web Kullanım Madenciliği Yöntemiyle Analizi: Fırat Üniversitesi Örneği. *e-Journal of New World Sciences Academy* 2008. 3(2).
- [2] Daş, R., İ. Türkoğlu, ve M. Poyraz, Web Kayıt Dosyalarından İlginç Örüntülerin Keşfedilmesi. *Fırat Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 2007. 19(4): s. 493-503.
- [3] Tuğ, E., M. Şakiroğlu, and A. Arslan, Auto-matic discovery of the sequential accesses from web log data files via a genetic algorithm. *Knowledge-Based Systems*, 2006. 19(3): p. 180-186.
- [4] Carmona, C.J., et al., Web usage mining to improve the design of an e-commerce website: OrOliveSur. com. *Expert Systems with Applications*, 2012. 39(12): p. 11243-11249.
- [5] Parvatikar, S. and B. Joshi. Analysis of user behavior through web usage mining. in *ICAST–International Conference on Advances in Science and Technology*. 2014. Citeseer.
- [6] Özkan, Y., Veri madenciliği yöntemleri. 2008: Papatya Yayıncılık Eğitim.
- [7] Silahtaroglu, G., Veri madenciliği. Papatya Yayınları, İstanbul, 2008.
- [8] Alpaydın, E., Zeki veri madenciliği: Ham veriden altın bilgiye ulaşma yöntemleri. Bilişim 2000 eğitim semineri, 2000.
- [9] Han, J. and M. Kamber, Data mining: concepts and techniques (the Morgan Kaufmann Series in data management systems). 2000.
- [10] Akpınar, H., Veri tabanlarında bilgi keşfi ve veri madenciliği. *İÜ İşletme Fakültesi Dergisi*, 2000. 29(1): p. 1-22.
- [11] Shannon, C., A Mathematical Theory of Communication. *The Bell System Technical Journal*, 1948. 27.

- [12] Quinlan, J.R., C4. 5: programs for machine learning. 2014: Elsevier.
- [13] Quinlan, J.R., Simplifying decision trees. International Journal of Man-Machine Studies, 1987. 27(3): p. 221-234.
- [14] Devasena, C.L. Comparative Analysis of Random Forest REP Tree and J48 Classifiers for Credit Risk Prediction. in International Conference on Communication, Computing and Information Technology (ICCCMIT-2014). 2015.
- [15] Kalmegh, S., Analysis of WEKA data mining algorithm REPTree, Simple CART and RandomTree for classification of Indian news. International Journal of Innovative Science, Engineering and Technology, 2015. 2(2): p. 438-46.
- [16] Srinivasan, D.B. and P. Mekala, Mining Social Networking Data for Classification Using REPTree. International Journal of Advance Research in Computer Science and Management Studies, 2014. 2(10).
- [17] Brownlee, J. Bagging and Random Forest Ensemble Algorithms for Machine Learning. 19.12.2017]; Available from: <https://machinelearningmastery.com/bagging-and-random-forest-ensemble-algorithms-for-machine-learning/>.
- [18] Breiman, L., Random forests. Machine learning, 2001. 45(1): p. 5-32.
- [19] Breiman, L., et al., Classification and regression trees. 1984: CRC press.
- [20] Logistic model tree. 19.12.2017]; Available from: https://en.wikipedia.org/wiki/Logistic_model_tree.
- [21] Landwehr, N., M. Hall, and E. Frank, Logistic model trees. Machine learning, 2005. 59(1-2): p. 161-205.
- [22] Parmaksız, M.Y., Google Analytics. Dike-yeksen Yayınevi, İstanbul, 2013.
- [23] Weka 3: Data Mining Software in Java. n.d. <http://www.cs.waikato.ac.nz/ml/weka/index.html>