



TV ve Set Üstü Cihaz Arayüzlerinin Kullanılabilirliğinin Değerlendirmesinde Makine Öğrenmesinin Kullanımı

Aycan Pekpazar^{1*}, Çiğdem Altın Gümüşsoy²

¹ İstanbul Teknik Üniversitesi, İşletme Fakültesi, Endüstri Mühendisliği Bölümü, İstanbul, Türkiye (ORCID: 0000-0001-9329-6936), kayaayca@itu.edu.tr

² İstanbul Teknik Üniversitesi, İşletme Fakültesi, Endüstri Mühendisliği Bölümü, İstanbul, Türkiye (ORCID: 0000-0003-2925-0954), altinci@itu.edu.tr

(3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications June 11-13, 2021)

(DOI: 10.31590/ejosat.946025)

ATIF/REFERENCE: Pekpazar, A. & Gümüşsoy Altın, Ç. (2021). TV ve Set Üstü Cihaz Arayüzlerinin Kullanılabilirliğinin Değerlendirmesinde Makine Öğrenmesinin Kullanımı. *Avrupa Bilim ve Teknoloji Dergisi*, (26), 41-46.

Öz

Teknoloji alanındaki hızlı gelişmelerle birlikte günümüzde geleneksel TV'ler birçok yeni özellik kazanarak akıllı TV'lere dönüşmüştür. Bu dönüşümle birlikte TV arayüzlerinin karmaşıklığı da giderek artmış ve kullanılabilirlik problemlerine sebep olmaya başlamıştır. Kullanılabilirlik problemlerinin ürün tasarımının erken aşamalarında belirlenmesi firmaların maliyetlerini düşürmekte ve müşterilere daha kullanılabilir sistemler sunulabilmektedir. Ancak kullanılabilirlik değerlendirmesinin çeşitli aşamaları uzman görüşlerine dayanmakta ve uzun süreler almaktadır. Makine öğrenmesi ve yapay zekâ teknolojileri pek çok alanda olduğu gibi kullanılabilirlik değerlendirmesinde de bazı süreçlerin otomasyonu ile süreçlerin hızlandırılması konusunda kullanılabilir. Bu çalışmanın amacı kullanılabilirlik problemlerinin öncelik düzeyleri açısından belirli örüntülere sahip olup olmadığını ilişkilendirme kuralları tekniği ile araştırmak ve çeşitli makine öğrenmesi algoritmaları (naive bayes, lojistik regresyon, hızlı geniş marjin, derin öğrenme, rastgele orman, gradyan artırma ağaçları, destek vektör makineleri teknikleri) yardımıyla kullanılabilirlik problemlerini önceliklerine göre sınıflandırmaktır. Bu amaçla Türkiye'nin önde gelen dijital platformlarından birisi olan Digitürk'ten TV ve set üstü cihaz arayüzünün yazılımcılar tarafından değerlendirilmesi sonucunda elde edilen 3695 problem temin edilmiştir. Elde edilen veri incelenerek toplamda kullanılabilirlikle ilgili 2752 problem belirlenmiştir. Analizler öncesinde metinlerden oluşan veri seti, sözcüklerine ayırma (tokenization), filtreleme, kök bulma (stemming) gibi ön işlemlerden geçirilerek analizler için hazır hale getirilmiştir. Çalışma kapsamında öncelik düzeyleri açısından kullanılabilirlik problemlerinin sahip olduğu örüntüler tespit edilmiştir. Ayrıca kullanılabilirlik problemleri farklı eğitim/test verisi oranları (50/50, 55/45, 60/40, 65/35, 70/30, 75/25, 80/20, 85/15, 90/10, 95/5) kullanılarak önceliklerine göre sınıflandırılmıştır. Sınıflandırma algoritmalarının performansları doğruluk oranı ve F1-skor metrikleri kullanılarak karşılaştırılmıştır. Çalışma sonucunda öncelik düzeylerine göre sınıflandırmada en yüksek doğruluk oranını (%76,21) destek vektör makineleri algoritması verirken en yüksek F1-skor değerini ise (%79,51) ile derin öğrenme algoritması vermiştir.

Anahtar Kelimeler: TV, Set üstü cihaz, Kullanılabilirlik Problemleri, Önceliklendirme, Makine öğrenmesi, Metin madenciliği.

The Use of Machine Learning in Evaluating the Usability of TV and Set-Top Boxes Interfaces

Abstract

With the rapid developments in technology, today traditional TVs have gained many new features and turned into smart TVs. With this transformation, the complexity of TV interfaces gradually has increased and started to cause usability problems. Identifying usability problems in the early stages of the product design reduces the costs of companies and provides more usable systems to customers. However, the various stages of usability evaluation are based on expert opinion and take a long time. Machine learning and artificial

* Sorumlu Yazar: İstanbul Teknik Üniversitesi, İşletme Fakültesi, Endüstri Mühendisliği Bölümü, İstanbul, Türkiye, ORCID: 0000-0001-9329-6936, kayaayca@itu.edu.tr

intelligence technologies can be used in usability evaluation as in many areas to automate and speed up the processes. This study aims to investigate whether usability problems have certain patterns in terms of importance levels using the association rules technique and to classify usability problems with the help of various machine learning algorithms (naive bayes, logistic regression, fast large margin, deep learning, random forest, gradient boosted trees, support vector machines) according to their importance levels. For this purpose, 3695 problems of a TV and set-top box interface determined by the software developers were obtained from Digiturk, which is one of Turkey's leading digital platforms. By examining the problems obtained, in total 2752 usability problems were determined. Before the analyze, the data set consisting of texts was made ready for analysis by pre-processing such as tokenization, filtering, stemming. As a result of the study, the patterns of usability problems according to their importance levels were obtained by using the association rules technique. Furthermore, usability problems were classified according to their priorities using different training/test data splitting ratios (50/50, 55/45, 60/40, 65/35, 70/30, 75/25, 80/20, 85/15, 90/10, 95/5). The performances of the classification algorithms were compared according to accuracy rate and F1-score metrics. As a result of the study, support vector machines had the highest accuracy level (76.21%) and deep learning algorithms had the highest F1-score (%79,51) in the classification of usability problems according to priority levels.

Keywords: TV, Set-top box, Usability, Machine learning, Text mining.

1. Giriş

Günümüzde hızlı teknolojik gelişmelerle birlikte evlerimizde kullandığımız geleneksel televizyonlar VOD (talep üzerine görüntü), internete bağlanma, oyun oynama, online alışveriş, sosyal medya platformlarına bağlanma gibi özellikler kazanarak giderek bilgisayarlara benzemeye başlamıştır. Her yeni eklenen özellikle birlikte TV arayüzlerinin karmaşıklığı da giderek artmış ve kullanılabilirlik problemlerine sebep olmaya başlamıştır. Kullanıcı ile doğrudan etkileşimi sağlayan kullanıcı arayüzlerinin kullanılabilirliği, geliştirilen yazılımın ve ürünün kalitesini belirleyen ve müşteri memnuniyetini (Lee vd., 2015; Dianat vd., 2019) etkileyen en önemli unsurlardan biridir.

Kullanılabilirlik, kullanıcı arayüzlerinin ne kadar kolay kullanılabilirliğini gösteren bir kalite özelliğidir (Nielsen, 1993). ISO tarafından ise kullanılabilirlik; "Belirli kullanıcıların belirli ortamlarda belirli görevleri etkili, verimli ve memnuniyet içerisinde gerçekleştirme seviyesi" olarak tanımlanmıştır (ISO, 9241-11,1998). Kullanılabilirlik kullanıcı testleri, sezgisel değerlendirme, anketler, bilişsel çözüm yolu gibi pek çok farklı yöntem kullanılarak değerlendirilmektedir. Ancak geleneksel kullanılabilirlik değerlendirmesi uzun zaman almakta ve çok fazla insan gücü gerektirmektedir (Ivory ve Hearst, 2001; Boza vd., 2014; El-Halees, 2014). Kullanılabilirlik değerlendirmesinin çeşitli süreçlerinin otomasyonu sayesinde kullanılabilirlik değerlendirmesinin süresi, maliyetleri ve uzman ihtiyacı azaltılabilir, değerlendirme süresi kısılacağı için daha fazla özelliğin ve alternatif tasarımın değerlendirilmesi mümkün olabilirken değerlendirmelerin tutarlılığı da artırılabilir (Ivory ve Hearst, 2001). Bu sürecin otomasyonu ile ilgili literatürde sınırlı sayıda da olsa çalışmalar yapılmıştır. Çalışmaların genel amacı çeşitli veri madenciliği tekniklerini kullanılabilirlik değerlendirmesinde kullanarak kullanılabilirlik problemlerindeki örüntüleri ve kullanılabilirlik faktörleri arasındaki ilişkileri bulmaktır.

González vd. (2006) web sitesi prototiplerinin kullanılabilirliğinin değerlendirilmesi için ilişkilendirilme kuralları tekniğini kullanmışlardır. Web sitesi prototipi ile ilgili 200 üzerinde kişinin katılımıyla bir anket çalışması yapmışlar ve sonuçları ilişkisel bir veri tabanında toplamışlardır. Anket sonuçları WEKA yazılımında Apriori algoritması kullanılarak analiz edilmiş ve kullanılabilirlik problemleri arasındaki sebep sonuç ilişkileri ortaya çıkarılmıştır.

González vd. (2008) kalitatif kullanılabilirlik test süreci için ilişkilendirme kuralları ve karar ağaçları tekniklerini kullanan yeni bir yaklaşım önermişlerdir. Önerdikleri yöntemin validasyonu için İspanyolca konuşan ülkelerdeki akademik web sayfalarının sezgisel değerlendirmesi sonucunda elde edilen 3450 kayıt WEKA platformu kullanılarak incelenmiş ve kullanılabilirlik problemlerinin sahip olduğu örüntüler belirlenmiştir.

Oztekin vd. (2013) e-öğrenme sistemlerinin kullanılabilirliklerinin değerlendirilmesi için makine öğrenmesi temelli bir değerlendirme sistemi geliştirmişlerdir. Makine öğrenmesi teknikleri ile birlikte çoklu lineer regresyon yöntemleri kullanılarak genel sistem kullanılabilirliğini etkileyen faktörler belirlenmeye çalışılmıştır. Ayrıca faktörlerin önem düzeylerini duyarlılık analizi ile tespit etmişler ve ciddiye endeksi geliştirmişlerdir. Çalışma kapsamında geliştirilen yöntemin etkinliğini test etmek üzere biyoloji dersi e-öğrenme uygulaması ile ilgili vaka çalışması yapılmıştır.

Boza vd. (2014) web kullanılabilirliği için ilişkilendirme kuralları ve karar ağaçları tekniklerini birlikte kullanmışlardır. Değişik web sayfalarının kullanılabilirlik uzmanları tarafından değerlendirmeleri sonucunda elde edilen 35 raporu analiz etmişlerdir. Web siteleri beş kullanılabilirlik faktörü (öğrenilebilirlik, etkililik, hatırlanabilirlik, hata önleme, memnuniyet) ve genel kullanılabilirlik düzeyi açısından uzmanlar tarafından değerlendirilmiştir. Weka yazılımı analizlerde kullanılmıştır. Analizler sonucunda kullanılabilirlik faktörleri ile ilgili kurallar elde edilmiştir.

El-Halees (2014) bir metin madenciliği tekniği olan görüş madenciliğini yazılım kullanılabilirliği değerlendirmesinde kullanmıştır. Kullanıcıların yazılımlar hakkındaki görüşlerini etkinlik, etkililik ve memnuniyet olmak üzere üç kullanılabilirlik faktörü doğrultusunda görüşlerini yazmaları istenmiştir. Toplanan 565 kullanıcı görüşü sırasıyla sözlük tabanlı, maksimum entropi ve kNN (en yakın k komşu) sınıflandırıcıları kullanılarak negatif ve pozitif olmak üzere sınıflandırılmıştır. Çalışma sonucunda ortalama %85,41 doğruluk oranı elde edilmiştir.

Wu vd. (2014) ürün kullanılabilirliğini değerlendirmek üzere web madenciliği tabanlı bir yaklaşım kullanmışlardır. Çevrimiçi müşteri yorumlarının hangilerinin kullanılabilirlikle ilişkili olduğunu belirlemek üzere apriori algoritması kullanılarak ilişkilendirme kuralları elde edilmiştir. Sonrasında yorumlar kullanılabilirlik problemi olma ve olmama durumuna göre

sınıflandırılmıştır. Ayrıca negatif ve pozitif yorum sayıları belirlenerek ürünlere ilişkin kullanılabilirlik değerlendirme sonuçları elde edilmiştir.

Etemudi vd. (2017), yazılım geliştirme sürecinde karşılaşılan kullanılabilirlik problemlerini belirlemek için Weka yazılımında apriori algoritması kullanılarak ilişkilendirme kuralları elde edilmiştir. Girdi olarak ise StackOverflow web sayfasında sorular-cevaplar kısmında Java, Python gibi programlama dilleri ile kullanıcıların sormuş olduğu 50000 soru incelenmiştir.

Sagar ve Saha (2017) kalitatif kullanılabilirlik değerlendirmesini woorank aracı ve veri madenciliği teknikleri (ilişkilendirme kuralları ve karar ağaçları) ile entegre ederek yeni bir yaklaşım geliştirmişlerdir. Geliştirdikleri yöntemin validasyonu için 50 tane akademik web sayfasının sezgisel değerlendirilmesine ilişkin 2475 kayıt incelenmiş ve kullanılabilirlik problemleri örüntüleri tespit edilmiştir.

Chamba-Eras vd. (2017), 24 üniversite web portalının kullanılabilirliği SIRIUS (system for evaluation of usability of web sites) sezgisel değerlendirme sistemi kullanılarak değerlendirilmiştir. Web portalları kullanılabilirlik düzeyleri açısından Rapidminer yazılımında kNN algoritması ve Öklidyen uzaklığı değerleri kullanılarak kümelere ayrılmıştır.

Literatürde incelenen çalışmalar genellikle kullanılabilirlik problemlerinin sahip olduğu örüntüleri belirlemek üzere ilişkilendirme kurallarını belirlemeye çalışmışlardır. Bu çalışma kapsamında ise literatürdeki diğer çalışmalardan farklı olarak kullanılabilirlik problemlerinin önem düzeyleri açısından sahip oldukları örüntüleri belirlemenin yanı sıra problemlerin önem düzeylerine göre sınıflandırılmaları da yapılmıştır. Ayrıca diğer çalışmalar büyük çoğunlukla web sitelerinin kullanılabilirliğinin değerlendirilmesinin veri madenciliği teknikleri ile otomasyonuna odaklanırken, bu çalışma diğerlerinden farklı olarak TV ve set üstü cihaz arayüzü kullanılabilirliği değerlendirmesi için veri madenciliği tekniklerini kullanmıştır. Bu özelliği ile de literatürdeki ilk çalışmadır.

Çalışmanın ikinci bölümünde çalışmanın metodolojisi, dördüncü bölümde ise yapılan uygulama çalışmasından bahsedilmiştir. Son bölümde ise çalışmanın sonuçları, kısıtları ve gelecek çalışma önerileri tartışılmıştır.

2. Materyal ve Metot

Çalışma kapsamında TV ve set üstü cihaz arayüzleri ile ilgili kullanılabilirlik problemleri veri madenciliği ve makine öğrenmesi teknikleri ile incelenmiştir. Kullanılabilirlik problemlerinin önem düzeyleri ile aralarında bulunan ilişkileri ve örüntüleri belirlemek üzere öncelikle ilişkilendirme kuralları kullanılmıştır. İkinci adımda ise kullanılabilirlik problemlerinin hangi öncelik düzeyine sahip olduğu sınıflandırma algoritmaları kullanılarak tahmin edilmiştir. Analizler içinse RapidMiner Studio 9.9 programı kullanılmıştır.

2.1. İlişkilendirme Kuralları

İlişkilendirme kuralları veri etiketlemeye gerek duyulmadan veriler arasındaki ilişkileri ve veri kümesindeki örüntüleri eğer/sonra şeklinde kurallar olarak belirlemeye yarayan gözetimsiz (unsupervised) bir veri madenciliği tekniğidir (Sagar ve Saha, 2017; Dökeroğlu vd.,2018). Bu yöntem genellikle pazar

sepeti analizi problemlerinin çözümünde kullanılmaktadır. İlişkilendirme kurallarının belirlenmesinde destek (support) ve güven (confidence) olmak üzere iki temel parametre bulunmaktadır. $(X \Rightarrow Y)$ kuralının destek değeri hem X hem Y'yi (XUY) içeren işlemler yüzdesi olarak tanımlanmaktadır. Diğer bir deyişle destek değeri, X ve Y nesnelere bir arada bulunma olasılıklarını göstermektedir. Bir kuralın güven değeri ise (XUY) 'yi içeren işlemlerin sayısının X'i içeren işlemler sayısına oranıdır (Srikant & Agrawal, 1995). Güven değeri tespit edilen ilişkinin kesinlik derecesini göstermektedir ve koşullu olasılık olarak tanımlanabilir (Han vd., 2011). İlişkilendirilme kurallarının üretilmesi için apriori, SETM ve FP-growth (frequent pattern growth) algoritmaları gibi birçok algoritma kullanılmaktadır. Bu çalışma kapsamında ilişkilendirme kurallarının üretilmesinde FP-growth algoritması kullanılmıştır. Ayrıca çalışma kapsamında minimum destek değeri 0,05; minimum güven değeri ise 0,75 olarak alınmıştır.

2.2. Sınıflandırma

Sınıflandırma, veri sınıflarını tanımlayan modelleri çıkarmak için kullanılan gözetimli bir veri analizi tekniğidir (Han vd, 2011). Sınıflandırma analizi, öğrenme ve sınıflandırma aşamalarından oluşur. Sınıflandırma modelleri, öğrenme aşamasında eğitim verisinde bulunan etiketleri belli olan verileri analiz eder ve verilerdeki özelliklere dayanarak sınıflara ilişkin modeller üretir ve sınıflandırma aşamasında bu modellere göre test verisindeki verilerin sınıf etiketlerini tahmin eder (Han vd., 2011). Bu çalışma kapsamında naive bayes, genelleştirilmiş lineer model, derin öğrenme, lojistik regresyon, hızlı geniş marjin, rastgele orman, gradyan artırma ağaçları ve destek vektör makinesi yöntemleri kullanılarak sınıflandırma modelleri oluşturulmuş ve sınıf etiketleri tahmin edilmiştir. Veriler rastgele olarak farklı eğitim/test verisi oranları (50/50, 55/45, 60/40, 65/35, 70/30, 75/25, 80/20, 85/15, 90/10, 95/5) kullanılarak ayrılmıştır. Her bir model 7 kez çalıştırılarak belirlenen performans metriklerinin ortalamaları ve standart sapmaları hesaplanmıştır.

2.2.1.Sınıflandırma Modelleri

Naive Bayes: Özelliklerin sınıf bilgisinden bağımsız olduğu varsayımıyla öğrenme sürecini basitleştirerek çalışan ve en olası sınıfı örneğe atayan bir sınıflandırma algoritmasıdır (Rish, 2001).

Genelleştirilmiş Lineer Model: Geleneksel doğrusal regresyon modellerinin bir uzantısı olan model, eğitim parametrelerinin maksimum olabilirlik değerlerine bağlı olarak yinelemeli ağırlıklandırılmış en küçük kareler yöntemini kullanarak tahmin yapar ve verinin normal, poisson, binom, gama vb. üstel dağılıma uydurulmasını sağlarlar (Nordin vd., 2020). Geleneksel doğrusal regresyonun aksine bağımlı değişkenin sayı veya oran olmadığı kategorik veya ikili değişken olduğu durumlarda kullanılabilirler (Faraway, 2016)

Derin Öğrenme: Çoğu derin öğrenme uygulaması, geri yayılım prosedürleri kullanılarak hesaplanan stokastik gradyan iniş (gradient descent) fonksiyonuyla eğitilen çok katmanlı ileri beslemeli sinir ağı yapılarını kullanmaktadır ve bir katmandan diğerine geçmek için önceki katmandan gelen girdilerin ağırlıklı toplamı hesaplanarak tanh veya rektifiye doğrusal birim (rectified linear unit, RELU) gibi doğrusal olmayan bir fonksiyondan geçirilmektedir (LeCun vd., 2015)

Lojistik Regresyon: Bir veya daha fazla bağımsız değişkene bağlı olarak kategorik (iki kategorili, aralarında doğal sıralama olan veya olmayan üç veya daha fazla kategorili) bağımlı

değişkenin tahmin edilmesi için kullanılan bir regresyon analizi türüdür (Menard, 2010). Lojistik regresyon, kategorik bağımlı değişken ile bir veya daha fazla sayıdaki bağımsız değişkenler arasındaki ilişkiyi bağımlı değişkenleri lojistik fonksiyon yardımıyla olasılık puanlarına dönüştürerek ölçümler.

Hızlı Geniş Marjin: Metin sınıflandırması gibi büyük çaplı sınıflandırma problemlerinin çözülmesi amacıyla Fan vd. (2008) tarafından geliştirilen hızlı marj öğrenicisini kullanan bir destek vektör makinesi türüdür.

Rastgele Orman: Tüm öznelikler kümesinden rastgele seçilen bir öznelik alt kümesi kullanan ağaçlardan oluşan ve yeni örneği çoğunluk oyuyla sınıflandıran bir yöntemdir (Oshiro vd., 2012).

Gradyan Artırma Ağaçları: Gradyan artırma modeli tahminleri bağımsız yapmak yerine kademeli olarak sırasıyla yaparak zayıf tahminleri güçlü tahminlere dönüştürmeyi amaçlayan regresyon veya sınıflandırma ağacı modellerinden oluşan bir topluluk algoritmasıdır (Hastie vd., 2009; Rapidminer, 2021).

Destek Vektör Makinesi: Temel amacı farklı sınıflara ait destek vektörleri arasındaki mesafeyi maksimize ederek sınıfları birbirinden ayıracak optimal düz çizginin yahut daha üst boyutlar için hiper-düzlemin bulunmasıdır (Suthaharan, 2016).

2.2.2. Performans Metrikleri

Sınıflandırma algoritmalarının performanslarının değerlendirilmesinde Tablo 1’de gösterilen karmaşıklık matrisi (confusion matrix) kullanılmıştır. Ayrıca algoritmaların hangisinin daha iyi sonuç verdiğini belirlemek üzere doğruluk oranı ve F1-Skor parametreleri kullanılmıştır. Doğruluk oranı, algoritma tarafından doğru sınıflandırılan test örneklerinin yüzdesi olarak tanımlanmaktadır (Kılınç vd., 2016). F1-Skor ise duyarlılık (precision) ve hassasiyet (recall) değerlerinin harmonik ortalamasıdır ve bu iki parametrenin etkisini birleştirmek için kullanılır.

Tablo 1. Karmaşıklık matrisi

Tahmin Edilen Sınıf		Gerçek Sınıf	
		Pozitif	Negatif
Pozitif	Doğru Pozitif (DP)	Yanlış pozitif (YP) (Tip I hata)	
	Yanlış negatif (YN) (Tip II hata)	Doğru negatif (DN)	
Negatif			

$$ACC = \frac{\sum \text{Doğru pozitif} + \sum \text{Doğru negatif}}{\sum \text{Toplam popülasyon}} = \frac{DP+DN}{DP+DN+YP+YN} \quad (1)$$

$$\text{Duyarlılık (precision)} = \frac{DP}{DP+YP} \quad (2)$$

$$\text{Hassasiyet (recall)} = \frac{DP}{DP+YN} \quad (3)$$

$$F1 - \text{Skor} = 2 \frac{(\text{Duyarlılık} * \text{Hassasiyet})}{(\text{Duyarlılık} + \text{Hassasiyet})} = \frac{2DP}{2DP+YP+YN} \quad (4)$$

3. Araştırma Sonuçları ve Tartışma

3.1. Veri

Türkiye’nin önde gelen dijital TV platformu olan Digiturk kendi TV ve set üstü cihaz arayüzlerini tasarlamaktadır. Çalışma kapsamında Digiturk’te geliştirilen arayüzlerle ilgili yazılımcılar yaptıkları testler sonucunda kullanılabilirlik problemleri de dâhil

olmak üzere pek çok fonksiyonel problem tespit etmekte ve önem düzeyleri ve problem detayları ile ilgili yorumlarını bir ortak çalışma platformuna girmektedirler. Çalışma kapsamında yazılımcıların belirlediği 3695 problem Digiturk’ten temin edilmiştir. Problemlerden kullanılabilirlikle ilgili olmayanlar ve benzer anlama gelecek şekilde yazılan tekrar eden problemler öncelikle veri setinden çıkarıldıktan sonra toplamda 2752 problem kalmıştır.

3.2. İlişkilendirilme Kuralları

Öncelikle metin verisi sözcüklere ayırma (tokenization), karakter dönüştürme, etkisiz sözcükleri filtreleme, uzunluğa göre sözcükleri filtreleme ve kök bulma ön işleme süreçlerinden geçirilmiştir. Daha sonra her sözcüğün bir kayıta bulunup bulunmaması durumunu gösteren 0-1 değerlerinden oluşan vektör dokümanı elde edilmiştir. Sonraki adımda FP-Growth algoritması kullanılarak her bir sözcüğün destek değerleri hesaplanmış ve ilişkilendirme kuralları elde edilmiştir.

İlişkilendirme modeli yüksek ve düşük önem düzeyine sahip kullanılabilirlik problemleri için sırasıyla çalıştırılmıştır. Yüksek ve düşük önceliğe sahip problemler için ilişkilendirilme kuralları Tablo 2’deki gibi çıkmıştır. Tablo 2’de gördüğümüz kuralları yorumlayacak olursak örneğin birinci satır için (occur ⇒ error) ilişki kuralının destek değeri %2,7, güven değeri ise %93 olarak bulunmuştur. Diğer bir deyişle, veri setinin %2,7’sinde “occur” ve “error” kelimeleri birlikte kullanılmıştır ve “occur” kökünden türetilen bir kelimenin bulunduğu kullanılabilirlik problemlerinin %93’ünde “error” kökünden türetilmiş en az bir kelime bulunmaktadır. Ayrıca (icon, display) ve (top, display) sözcüklerini içeren problemlerin düşük önceliğe sahip problemler olduğu gözlemlenmiştir.

Tablo 2. Öncelik düzeylerine göre problemler için ilişkilendirme kuralları

No	Öncül	Sonuç	Destek	Güven
Yüksek	occur	error	0,027	0,93
	black	screen	0,027	0,93
	temporary	error	0,027	0,95
	occur, temporary	error	0,024	0,97
Düşük	icon	display	0,051	0,79
	top	display	0,020	0,79

3.3. Sınıflandırma

TV ve set üstü cihaz arayüzünün kullanılabilirliğini önemli ölçüde etkilediği için yazılımcıların düşük öncelikli problemleri çözmek yerine öncelikle yüksek önceliğe sahip olan problemleri çözmesi gerekmektedir. Bu sebeple TV ve set üstü cihaz arayüzünü test eden yazılımcıların kullanılabilirlik problemleri için yüksek ve düşük olmak üzere tanımladıkları öncelik düzeyleri kullanılarak iki sınıflı bir model oluşturulmuştur. Kullanılabilirlik problemlerinin %52,58’i yüksek önem düzeyine sahipken %47,42’si ise düşük önem düzeyine sahiptir. Farklı eğitim/test verisi oranları (50/50, 55/45, 60/40, 65/35, 70/30, 75/25, 80/20, 85/15, 90/10, 95/5) kullanılarak oluşturulan veri setleri ile ilgili sınıflandırma algoritmaları çözülmüştür. Her bir algoritma için en iyi doğruluk oranı ve F1-skor değerini veren modelin sonucu seçilerek algoritmaların çözüm performanslarını gösteren Tablo 3 oluşturulmuştur. Doğruluk oranları %59,87 ile %76,21 değerleri arasında değişirken, F1- skor değerleri ise %69,17 ile %79,51 değerleri arasında değişmektedir. En iyi doğruluk oranını (%76,21) destek vektör makineleri algoritması verirken en yüksek F1-skor değerini ise (%79,51) ile derin öğrenme algoritması vermiştir. Ancak her iki performans

metriğinin de standart sapması diğer algoritmalarla kıyasla daha yüksektir. Diğer algoritmaların sonuçları standart sapmalar dikkate alınarak incelendiğinde hızlı geniş marjin ve gradyan artırma ağaçları algoritmalarının standart sapma değerleri düşük daha genellenebilir sonuçlar verdiği söylenebilir. Genel olarak bu doğruluk oranları ve F1-skor değerleri çok yüksek olmamasına rağmen arayüz değerlendirme sürecinin karmaşıklığı ve sübjektifliği düşünüldüğünde makine öğrenmesi tekniklerinin bu konuda gelecek vadettiği söylenebilir.

4. Sonuç

Bu çalışma kapsamında TV ve set üstü cihaz arayüzü ile ilgili kullanılabilirlik problemleri ilişkilendirme kuralları tekniği ile incelenerek önem düzeyleri açısından sahip oldukları örüntüler belirlenmiştir. Ayrıca kullanılabilirlik problemleri naive bayes, logistik regresyon, hızlı geniş marjin, derin öğrenme, rastgele orman, gradyan artırma ağaçları, destek vektör makineleri teknikleri ile önceliklerine göre sınıflandırılmıştır. Çalışma sonucunda kullanılabilirlik problemlerinin çözülmesi için öncelikle hangi problemlerin ele alınması gerektiğini belirleme konusunda makine öğrenmesi teknikleri %76,21 doğruluk oranı ve %79,51 F1-skoru düzeyinde bir performans sergilemiştir. Sonuçlara göre makine öğrenmesi teknikleri kullanılabilirlik problemlerinin önceliklerini tahmin etme konusunda umut vadeden sonuçlar elde etmiştir. Bu sayede tespit edilen kullanılabilirlik problemlerinin önceliklerinin belirlenmesi kolaylaşacaktır, yazılımcılar öncelikle önem düzeyi yüksek problemlere odaklanabilecek böylece sistem kullanılabilirliği hızlı bir şekilde yükseltilebilecektir.

Çalışmanın bazı kısıtları da bulunmaktadır. Bu çalışmada sadece bir TV ve set üstü cihaz arayüzünün geliştirilmesi ile ilgili problemler toplanmıştır. Gelecekte farklı TV ve set üstü cihaz arayüzlerinden toplanan verilerle de benzer çalışmalar yapılması daha genellenebilir sonuçlar elde edilmesini sağlayacaktır. Ayrıca daha çok veri toplanması sayesinde daha güvenilir örüntüler elde edilebileceği gibi sınıflandırma algoritmaları da daha iyi performans sergileyebilecektir.

5. Teşekkür

Bu çalışma Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından TÜBİTAK 3001 programı ile desteklenmiştir (Proje numarası: 217M143, 2018). TÜBİTAK'a katkılarından dolayı teşekkür ederiz.

Kaynakça

Boza, B. C., Schiaffino, S., Teyseyre, A., & Godoy, D. (2014). An approach for knowledge discovery in a web usability context. *In Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems*, 393-396.

Chamba-Eras, L., Jacome-Galarza, L., Guaman-Quinche, R., Coronel-Romero, E., & Labanda-Jaramillo, M. (2017, April). Analysis of usability of universities Web portals using the Prometheus tool-SIRIUS. *In 2017 Fourth International Conference on eDemocracy & eGovernment (ICEDEG)*, IEEE, 195-199.

Dianat, I., Adeli, P., Jafarabadi, M. A., & Karimi, M. A. (2019). User-centred web design, usability and user satisfaction: The case of online banking websites in Iran. *Applied ergonomics*, 81, 102892.

Dökeroğlu, T., Malık, Z. M. M., & Shadi, A. S. (2018). Gözetimsiz Makine Öğrenme Teknikleri ile Miktarla Dayalı Negatif Birlikte Kural Madenciliği. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 6(4), 1119-1138.

El-Halees, A. M. (2014). Software Usability Evaluation Using Opinion Mining. *JSW*, 9(2), 343-349.

Etamadı, V., Bushehrian, O., & Akbari, R. (2017). Association rule mining for finding usability problem patterns: A case study on StackOverflow. *In 2017 International Symposium on Computer Science and Software Engineering Conference (CSSE)*, IEEE, 24-29.

Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R., & Lin, C. J. (2008). LIBLINEAR: A library for large linear classification. *Journal of machine learning research*, 9(Aug), 1871-1874.

Faraway, J. J. (2016). Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. CRC press.

González, M. P., Granollers, T., & Lorés, J. (2006). A hybrid approach for modelling early prototype evaluation under user-centred design through association rules. *In International Workshop on Design, Specification, and Verification of Interactive Systems*, Springer, Berlin, Heidelberg, 213-219.

González, M. P., Lorés, J., & Granollers, A. (2008). Enhancing usability testing through datamining techniques: A novel approach to detecting usability problem patterns for a context of use. *Information and software technology*, 50(6), 547-568.

Han, J., Kamber, M., & Pei, J. (2011). Data mining concepts and techniques third edition. *The Morgan Kaufmann Series in Data Management Systems*, 83-124.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. *Springer Science & Business Media*.

ISO 9241-11 1998. Ergonomic requirements for office work with visual display terminals (VDTs). Part 11: Guidance on usability.

Ivory, M. Y., & Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys (CSUR)*, 33(4), 470-516.

Kılınc, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimşek, M., & Özçift, A. (2016). KNN algoritması ve r dili ile metin madenciliği kullanılarak bilimsel makale tasnifi. *Marmara Fen Bilimleri Dergisi*, 28(3), 89-94.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>

Lee, D., Moon, J., Kim, Y. J., & Mun, Y. Y. (2015). Antecedents and consequences of mobile phone usability: Linking simplicity and interactivity to satisfaction, trust, and brand loyalty. *Information & Management*, 52(3), 295-304.

Menard, S. (2010). Logistic regression: From introductory to advanced concepts and applications. Sage.

Nielsen, J. (1993). Usability Engineering, Academic Press.

Nordin, N. D., Zan, M. S. D., & Abdullah, F. (2020). Generalized linear model for enhancing the temperature measurement performance in Brillouin optical time domain analysis fiber sensor. *Optical Fiber Technology*, 58, 102298.

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest? *In International workshop on machine learning and data mining in pattern recognition*, Springer, Berlin, Heidelberg, 154-168.

- Oztekin, A., Delen, D., Turkyilmaz, A., & Zaim, S. (2013). A machine learning-based usability evaluation method for eLearning systems. *Decision Support Systems*, 56, 63-73.
- Rapidminer. (2021). Gradient Boosted Trees. Retrieved from: https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/gradient_boosted_trees.html
- Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 3(22), 41-46.
- Sagar, K., & Saha, A. (2016). Enhancing usability inspection through data-mining techniques: an automated approach for detecting usability problem patterns of academic websites. In *International Conference on Intelligent Human Computer Interaction*, Springer, Cham, 229-247.
- Srikant, R., & Agrawal, R. (1995). Mining generalized association rules. In *21st VLDB Conference Zurich*, Switzerland, 407-419.
- Suthaharan S. (2016) Modeling and Algorithms. In: Machine Learning Models and Algorithms for Big Data Classification. *Integrated Series in Information Systems*, Springer, Boston, MA, 123-143.
- Wu, M., Wang, L., Li, M., & Long, H. (2014). An approach of product usability evaluation based on Web mining in feature fatigue analysis. *Computers & Industrial Engineering*, 75, 230-238.