



## Bir Öğrenci Bilgi Sisteminin Kullanılabilirliğinin Makine Öğrenmesi Teknikleriyle Tahmin Edilmesi

Denizhan DEMİRKOL<sup>1\*</sup>, Elif KARTAL<sup>2</sup>, Çağla ŞENELER<sup>3</sup>, Sevinç GÜLSEÇEN<sup>2</sup>

<sup>1</sup>Aydın Adnan Menderes Üniversitesi, Söke İşletme Fakültesi, Yönetim Bilişim Sistemleri Bölümü, İstanbul

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

<sup>3</sup>Yeditepe Üniversitesi, Ticari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, İstanbul

### Özet

Sistem kullanılabilirliği, bir sistemin özellikle tasarım ve test aşamalarında odaklanması gereken unsurlardan biridir, çünkü sistemin daha iyi hale getirilmesi için sistem yöneticilerine geri bildirim sağlamaktadır. Literatürde, sistem kullanılabilirliğinin değerlendirilmesi için Sistem Kullanılabilirlik Ölçeği (*System Usability Scale-SUS*) altın standart yöntem olarak yaygın şekilde kullanılmaktadır. Bunun yanı sıra günümüzde yapay zekânın alt çalışma alanlarından biri olan makine öğrenmesi de sistem kullanılabilirliğinin değerlendirilmesi konusunda araştırmacılara yeni ufuklar sağlamaktadır. Bu çalışmada, bir Öğrenci Bilgi Sisteminin (ÖBS) kullanılabilirliğinin makine öğrenmesi teknikleriyle tahmin edilmesi hedeflenmiştir. Çalışma yönteminde; Veri Madenciliği için Çapraz Endüstri Standard Süreç Modeli (*CRISP-DM*) kullanılmıştır. Analizler; Türkiye'deki bir vakıf üniversitesine ait bir ÖBS'yi kullanan 324 öğrencinin SUS'un Türkçe versiyonuna (SUS-TR) verdiği yanıtların bulunduğu "sus1" adlı veri seti ile öğrencilerin yaş, cinsiyet, öğrenim gördüğü bölüm eklenerek oluşturulan "sus0" adlı veri seti üzerinde gerçekleştirilmiştir. C4.5 Karar Ağacı Algoritması, Basit (*Naive*) Bayes Sınıflandırıcı ve k-En Yakın Komşu Algoritması ile farklı modeller kurularak modellerin performans değerlendirilmesi yapılmıştır. %80'e %20'lik Hold-out ayrımıyla gerçekleştirilen analizlerde en iyi performans, k-En Yakın Komşu Algoritmasıyla "sus0" veri seti üzerinde elde edilmiştir (k=20 için doğruluk = 0.698, F-ölçütü = 0.796).

**Anahtar Kelimeler:** Danışmanlı Öğrenme, Kullanılabilirlik, Öğrenci Bilgi Sistemi, Sistem Kullanılabilirlik Ölçeği

### Predicting Usability of a Student Information System by Using Machine Learning Techniques

#### Abstract

System usability is one of the key elements that should be focused on, especially during the design and test phases of a system, because it provides feedback to system administrators in order to improve the system. In the literature, System Usability Scale (SUS) is widely used as the gold standard method to evaluate system usability. Today, machine learning, which is one of the subfields of artificial intelligence, also provide new perspectives on the evaluation of system usability. In this study, it is aimed to predict usability of a Student Information System (SIS)

by using machine learning techniques. In the study method, the Cross-Industry Standard Process for Data Mining (CRISP-DM) steps have been followed. Analysis are performed on two different datasets namely "sus0" and "sus1". "sus0" dataset is consisted of demographic characteristics (age, gender, department) of 324 students using a SIS of a foundation university in Turkey, also their responses to the Turkish version of the SUS (SUS-TR). "sus1" includes only responses to the SUS-TR. C4.5 Decision Tree Algorithm, Naive Bayes Classifier and k-Nearest Neighbor Algorithm are used to create models and performance of the models are evaluated. In the analysis with 80% to 20% hold-out method, the best performance was obtained on the "sus0" data set with k-Nearest Neighbor Algorithm (accuracy = 0.698, F-measure = 0.796 for k = 20).

**Keywords:** Supervised Learning, Usability, Student Information System, System Usability Scale

## 1 Giriş

Uluslararası Standartlar Örgütü (*International Organization for Standardization - ISO*) tarafından belirlenen ISO 9241-11: 1998 [1] numaralı standartta göre kullanılabilirlik; bir ürünün, belirli bir kullanıcı grubu tarafından, belirlenen bağlam ve amaçlar çerçevesinde etkililiği, verimliliği ve kullanıcı memnuniyeti olarak tanımlanmaktadır. Bir hedefin ya da görevin elde edilme derecesi etkililik, bir hedefe ulaşmak için gereken çaba miktarı verimlilik, bir ürünün kullanılırken hissedilen rahatlık seviyesi ve ürünün bir araç olarak kullanılırken ne kadar kabul edilebilir olduğu ise kullanıcı memnuniyeti olarak ifade edilmektedir [2]. Nielsen [3] kullanılabilirliği; öğrenilebilirlik, verimlilik, hatırlanabilirlik, hatalar ve memnuniyet olmak üzere başlıca beş başlık altında incelemiştir:

- Öğrenilebilirlik, kullanıcıların arayüzle etkileşime geçtiği ilk zamanda temel görevleri ne kadar kolaylıkla başarabildikleridir.
- Verimlilik, kullanıcıların tasarımı öğrendikten sonra görevleri ne kadar hızlı sürede yerine getirdikleridir.
- Hatırlanabilirlik, kullanıcıların sistemi belirli bir süre kullanıma ara verdikten sonra ne kadar kolaylıkla yeniden yeterli bir şekilde kullanabildikleridir.
- Hatalar, kullanıcıların sistemi kullanırken ne kadar hata yaptıkları, bu hataların ne kadar önemli olduğu ve bu hataları ne kadar kolay bir şekilde telâfi ettikleridir.
- Memnuniyet, kullanıcıların sistemi kullanırken hissettikleri güzellik ölçütüdür.

Günümüzde her bir web sitesinin veya sistemin çok fazla sayıda alternatifi olduğundan bir sistemden diğerine kolaylıkla geçilebilmektedir. Nielsen [4] bir web sitesinin kullanılabilir olmadığı durumlarda insanların sistemden ayrıldıklarını belirtmiştir. Bu da kullanılabilir sistemler geliştirmenin önemini vurgulamaktadır.

Hornbæk [5], kullanılabilirliğin doğrudan ölçülemeyeceğini ve uygun kullanılabilirlik değerlendirme yöntemi seçiminin zorlayıcı olduğunu belirtmiştir. Sweeney, Maguire ve Shackel [6]; uzman, teori ve kullanıcı temelli olmak üzere üç farklı kullanılabilirlik değerlendirme metodunun kullanıcıların sistemi ne kadar iyi öğrendiklerini ve kullandıklarını anlamak için tercih edilebileceğini belirtmiştir. Uzman temelli yaklaşımın, kullanılabilirlik uzmanlarının bulunmasının zor ve maliyetli olması gibi birçok dezavantajı vardır [7]. Kullanıcı temelli yaklaşımın ise, kullanılabilirlik testlerinin her zaman yapay bir test ortamı olması, test sonuçlarının geliştirilen ürünün başarılı olacağını garantisini vermemesi, kullanılabilirlik testi için seçilen katılımcıların nadiren gerekli hedef kullanıcı grubunu yansıtmaması ve kullanılabilirlik testinin her zaman uygulanan en iyi teknik olmaması gibi sınırlılıkları vardır [8]. Bunlardan farklı olarak model temelli yaklaşımda [9], kullanıcının varsayılan sistemi elde etmek için planlanan sistemi nasıl kullanacağını bir model aracılığı ile belirtmektedir. Bu yaklaşım ile oluşturulan tahminlere bağlı modeller, kullanılabilirlik testi ile elde edilen deneysel ölçümlerin yerini alabilir veya bu ölçümleri tamamlayabilir. Ayrıca model, kullanıcı görevi ile sistem tasarımı arasındaki ilişki hakkında yararlı bilgileri de içerir.

Sağlıkta hastane takip sistemleri, turizmde otel rezervasyon sistemleri, bankacılık ve sigortacılık için müşteri takip sistemleri ve bu sistemlerin kullanılabilirliği hem hizmet veren hem de hizmet alan için ne kadar önemliyse, eğitimde de Öğrenci Bilgi Sistemleri (ÖBS'ler) ve ÖBS'lerin kullanılabilirliği eğitim personeli, idari personel ve öğrenciler için aynı öneme sahiptir. ÖBS'ler geçmiş, mevcut ve gelecek öğrencilerin kayıt yönetimini sağlama, öğrencilerin başvuru, kayıt, geçmiş/güncel akademik performansını kapsama, faturalandırma ve ücret yönetimini gerçekleştirme, mezuniyet ve sınavlar gibi önemli olayları işleme, sınıf programını yapma, operasyonel ve yönetimle ilgili konuları bildirme ve ilişkili verileri diğer pek çok uzman sisteme besleme gibi birçok işleve sahiptir [10]. Geçtiğimiz yıllarda online eğitimin de giderek önem kazanmasıyla ÖBS'lerin de popülerliği giderek artmış ve birçok eğitim kurumu bu sistemleri veri yönetiminde kullanmaya başlamıştır. Literatür incelendiğinde ÖBS kullanılabilirliği üzerine gerçekleştirilen çalışmaların fazla sayıda olmadığı gözlemlenmiştir. Oysaki bir ÖBS'nin benimsenmesindeki en büyük engellerden birisi de kullanılabilirlik sorunları olarak ifade edilmektedir [11].

ÖBS'ler sadece yükseköğretimde değil kamu ve özel sektörde eğitimle ilgili neredeyse her kurumda öğrenci, eğitmen ve idari personel olmak üzere üç temel kullanıcı grubu tarafından çok çeşitli fonksiyonlara sahip bir bilgi sistemidir. Ayrıca etkili ve kullanılabilir bir ÖBS sadece personel ve öğrencileri tatmin etmekle kalmayacak, aynı zamanda akademik iş akışının sürdürülebilirliğini de sağlayacaktır. İlkokul, ortaokul, lise ve üniversitelerde iyi yapılandırılmış ÖBS'lere ihtiyaç duyulmakta, ÖBS'lerle ilgili yapılan araştırmalar bu sistemlerin daha üretken hâle getirilmesi için üzerinde çalışılması gerektiğini göstermektedir. Ayrıca, literatürde her ne kadar sistem kullanılabilirliği belirli ölçekler yardımı ile değerlendirilse de, Karat [12] bir ölçek için uygun test katılımcılarının seçimini ve ölçek değerlendirme süreçlerini genel olarak maliyet gerektiren zorlu birer görev olduğunu belirtmiştir. Bu noktadan hareketle; eğitim kurumları için oldukça önemli olan ÖBS'lerin kullanılabilirlik durumlarının, yine bu sistemin kullanıcılarının verisine dayalı elde edilen modeller yardımıyla tahmin edilmesinin daha hızlı ve etkili bir değerlendirme yöntemi olabileceği

düşünülmüştür. Bu kapsamda; bu çalışmada Türkiye'deki bir vakıf üniversitesine ait bir ÖBS'nin kullanılabilirliğinin makine öğrenmesi tekniklerinden faydalanarak tahmin edilmesi amaçlanmıştır. Analizlerde; öğrencilerin orijinal versiyonu Brooke [13] tarafından geliştirilen Sistem Kullanılabilirlik Ölçeği'nin (*System Usability Scale-SUS*), Demirkol [14] tarafından geçerlilik/güvenilirlik çalışmaları gerçekleştirilen ve Türkçeye çevirisi yapılan Sistem Kullanılabilirlik Ölçeğinin Türkçe versiyonuna (SUS-TR) verdikleri cevaplar kullanılmıştır. Ayrıca öğrencilerin yaş, cinsiyet ve öğrenim gördüğü bölüm bilgisi de yine analizlerde kullanılmıştır. Yapay zekânın bir alt dalı olan makine öğrenmesi teknikleriyle gerçekleştirilen bu çalışmanın ÖBS kullanılabilirlik çalışmalarına ve ileride konuyla ilgili yapılacak araştırmalara ışık tutacağına inanılmaktadır.

## 2 Literatür taraması

Korvald, Kim ve Reza [15] sayfaların düzenlenmesiyle ilgili kullanılabilirlik sorunlarını ortaya çıkarmak için makine öğrenmesi tekniklerini kullanmış, bir web sitesinin kullanılabilirlik testi için bilgi mimarisi adını verdikleri otomatik bir yöntem önerisi sunmuştur. Çalışma; destek vektör sınıflandırıcıları, rasgele orman, karar ağacı, regresyon gibi modelleri içermektedir. Model performans değerlendirmesi farklı büyüklükteki veri setleri üzerinde 10-kat çapraz geçirme yöntemi kullanarak yapılmış ve mümkün olan en iyi modelin belirlenmesi amaçlanmıştır. En tutarlı model, standart sapması 0.02826 olan ve 50-li ve 100-lü veri setleriyle gerçekleştirilen doğrusal regresyon modeli olmuştur. Bununla birlikte çalışmada rastgele orman sınıflandırıcısı, veri setindeki önemli niteliklerin tanımlanmasında başarılı olmuştur. Bu modelden elde edilen nitelikler, bir site sahibine bilgi mimarisinin ilk olarak hangi yönlerinin geliştirilmesi gerektiğini önermek için kullanılabilirliği belirtilmiştir.

Oztekin [16] bir vakıf üniversitesinin ÖBS'sini kullanarak yaptığı çalışmada, kullanılabilirlik değerlendirmesi ve web tabanlı bilgi sistemi tasarımı için bir karar destek sistemi önermiştir. Araştırmada web tabanlı bilgi sisteminin genel kullanılabilirliği ile kullanılabilirliği belirleyici faktörler arasındaki temel ilişkiler; destek vektör makineleri, yapay sinir ağları ve karar ağaçları gibi üç makine öğrenmesi yöntemi ve istatistiksel bir teknik olan çoklu doğrusal regresyon ile ortaya

çıkarılmaya çalışılmıştır. Çalışma sonuçlarına göre önerilen karar destek sisteminin, kullanılabilirlik problemlerinin belirlenmesi ve öğrenci bilgi sisteminin kullanılabilirliğinin ve niteliğinin artırılması için kullanılabileceği belirtilmiştir.

Bu alanda yapılan çalışmalara; Oztekin, Delen, Turkyilmaz ve Zaim [17] tarafından e-öğrenme sistemlerinin kullanılabilirliğinin makine öğrenmesi teknikleriyle tahmini üzerine gerçekleştirdiği çalışma bir başka bir örnek olarak verilebilir. Yazarlar; bir öğrenme yönetim sistemi olan Moodle'nin kullanılabilirliği ve tahminde kullanılan nitelikler arasındaki temel ilişkiyi veren tahmin modellerini geliştirmek için yukarıdaki çalışmada [16] bahsedilen yöntemlerden faydalanmıştır. Çalışmada önerilen yöntemin, bir e-öğrenme sisteminin hedeflenen son kullanıcı grubu için kullanılabilirliğini en üst düzeye çıkarmak amacıyla iyileştirilmelerin yapılması gerektiği konusunda kullanılabilirlik uzmanlarına rehberlik sağladığı belirtilmektedir [17].

Davis [18] yaptığı çalışmada, web siteleri için kullanılabilirlik problemlerini tanımlamaya yönelik yeni bir yaklaşım sunmuştur. Bu yaklaşımda, bir sitenin çeşitli özelliklerinin kullanıcılar için ne derecede önemli olduğu tanımlanırken aynı zamanda web sayfalarının özellikleri kullanılabilirlik sorunları açısından ortaya çıkarılmaktadır. Çalışmada doğrusal regresyon modellerinden faydalanılmıştır. Sonuç olarak web sitesi yöneticilerinin, kullanılabilirlik denetimine ihtiyaç duyulan web sayfalarını test ederken kullanılabilirlik sorunlarını açığa çıkarmak için bu yaklaşımdan yararlanabilecekleri gösterilmiştir.

Horng, Kittur, Hong ve Faloutsos [19] yaptıkları araştırmada, fazla sayıda ve birbiriyle ilişkili verinin kullanıcılar tarafından etkileşimli olarak keşfedilmesini sağlamak için, görselleştirme ve makine öğrenmesini birleştiren karma bir yaklaşımı kullanan Apolo sistemini tanıtmışlardır. Apolo'nun kullanılabilirliği de on iki katılımcıya, mevcut bir araştırma belgesini güncellemek için ilgili yeni makaleleri bulma görevi verilerek değerlendirilmiştir. Araştırmada katılımcılar arasında iki koşuldan oluşan bir deney tasarımı kullanılmıştır. Bu iki koşul; katılımcıların belgeleri araştırmak için Apolo'yu veya Google Scholar'ı kullanmasıdır. Değerlendirme sonuçları göstermektedir ki, Apolo'yu kullanan katılımcılar verilen görev için Google Scholar'ı kullanan

katılımcılara göre daha ilgili makaleler bulmuşlardır.

Maragoudakis, Tselios, Fakotakis ve Avouris [20] yaptıkları çalışmada, bir kısa mesaj (SMS) kullanılabilirliğini Bayes ağlarını kullanarak arttırmayı hedeflemiş, bir SMS'in derlenme sürecinde Yunan harflerinin oluşumunu tahmin eden bir algoritma sunmuştur. Geliştirilen ağ modeli; kullanıcının verilen bir kelimenin belirli bir harfinin olasılığını bir, iki veya üç harf öncesinde %95 kesinlikle çıkarabilmektedir. Önerilen yöntemle kelime düzenlemede geleneksel yöntemlere göre %34.72 oranında daha fazla iyileştirme elde edilmiştir.

### 3 Materyal ve metod

Çalışma yönteminde; Veri Madenciliği için Çapraz Endüstri Standard Süreç Modeli (*Cross-Industry Standard Process for Data Mining – CRISP-DM*) [21] adımları izlenmiştir. CRISP – DM; problemin tanımlanması (*business understanding*), veriyi anlama (*data understanding*), veri hazırlama (*data preparation*), model kurma (*modeling*), model değerlendirme ve seçimi (*evaluation*) ile modelin uygulamaya geçirilmesi (*deployment*) adımlarından oluşmaktadır [21,22]. Bu çalışmanın giriş ve literatür taraması bölümlerinde, çalışmaya konu olan problem detaylı bir biçimde ele alındığından yöntem bölümü veriyi anlama adımı ile başlamıştır.

#### 3.1 Veriyi anlama

CRISP-DM kapsamında; Türkiye'deki bir vakıf üniversitesine ait bir ÖBS'yi kullanan 324 öğrencinin yaş, cinsiyet, bölüm bilgileri ile SUS-TR'ye verdikleri yanıtlarla ve SUS-TR skorlarına ait bir veri seti kullanılmıştır [14]. Veri setinde yalnızca yaş değişkeni nümerik olup, cinsiyet (kadın/erkek) ve bölüm (bilgisayar teknolojileri bölümlerinde okuyan öğrenci grubu ve diğer grup olmak üzere) nitelikleri ikili kategorik biçimdedir. Öğrencilerin okudukları bölümün bu şekilde ikiye ayrılmasındaki neden, bilgisayarla ilgili bir bölümde öğrenim görmenin sistem kullanılabilirliği kararı üzerine etkisinin daha ayırt edici olduğu düşüncesidir [14]. SUS-TR'de yer alan soruların yanıtları 1 (kesinlikle katılmıyorum) – 5 (kesinlikle katılıyorum) olmak üzere 5-li likert biçimdedir. Makine öğrenmesi analizlerinde kolaylık sağlama için SUS-TR'deki 10 madde, her maddenin ana temasına bağlı kalınarak, "kullanma sıklığı, karmaşıklık, kolaylık, teknik desteğe ihtiyaç, tutarsızlık, entegrasyon, öğrenilebilirlik,

elverişsizlik, kendinden eminlik ve ön bilgi gerektirme” biçiminde kısaltılarak kullanılmıştır.

Ölçeğin standart hesaplama yöntemine göre, sistem kullanılabilirlik durumunu yansıtan SUS puanlarını - çalışmadaki adıyla “karar” niteliği - katılımcıların ölçeğe verdikleri yanıtlara göre 100 puan üzerinden hesaplanmaktadır. Puan değeri arttıkça algılanan kullanılabilirlik de artmaktadır [23]. Makine öğrenmesi analizlerinde; ÖBS'nin kullanılabilirlik durumunu belirten “karar” niteliği hedef nitelik, “karar” niteliği dışındakiler ise tahmini sağlayan nitelikler olarak kullanılmıştır.

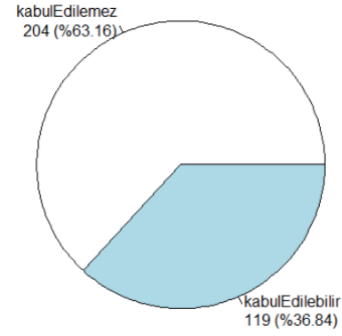
### 3.2 Veri hazırlama

Analizlerde hedef nitelik olarak belirlenen “karar” niteliği SUS-TR puanlarını tutan nümerik bir niteliktir; ancak bu çalışmada “Model Kurma” bölümünde verilen danışmanlı öğrenme algoritmalarının kullanılabilmesi için Bangor, Kortum ve Miller [24] tarafından yapılan çalışmadan faydalanılarak öğrencilerin ÖBS'nin kullanılabilirlik durumu kararını tutan hedef nitelik kategorik hâle getirmiştir. Bangor ve diğerleri [24] tarafından 1000 kişiyle yapılan çalışmada, SUS skorlarını kategorize etmek amacıyla 10 maddelik SUS ölçeğine ek bir soru eklenmiştir. Sistem kullanılabilirliği; “hayal edilebilenin en kötüsü”, “berbat”, “zayıf”, “geçer”, “iyi”, “mükemmel” ve “hayal edilebilenin en iyisi” biçiminde değerlendirilmiştir. Çalışma sonuçlarına göre; katılımcıların SUS puanları ile ek soruya verilen cevaplar arasında yüksek derecede ilişki olduğu tespit edilmiş ( $\alpha < 0.01$ ,  $r = 0.822$ ), sistem kullanılabilirliği için “geçer”e karşılık gelen SUS puan ortalaması ise 50.9 ( $\sigma = 13.8$ ) biçiminde bulunmuştur.

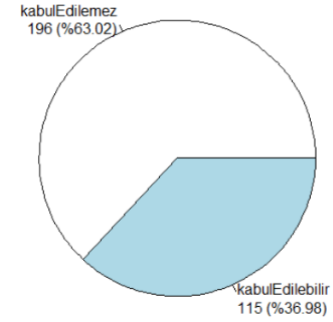
Bu nedenle bu çalışmada, veri setindeki hedef nitelik “kabul edilebilir” ve “kabul edilemez” biçiminde ikiye ayrılırken Bangor ve diğerleri [24] tarafından gerçekleştirilen bu kapsamlı çalışma göz önünde bulundurulmuş, 50.9'un üzerinde alınabilecek en küçük SUS skoru 52.5 kesme noktası olarak seçilmiştir. Böylelikle, kullanılabilirlik “karar” puanı 52.5'in altında olan katılımcıların sistemin kullanılabilirliği üzerine verdiği nihai karar “kabul edilemez”, 52.5 ve üzerinde puana sahip katılımcıların sistemin kullanılabilirliği üzerine verdiği nihai karar ise “kabul edilebilir” kategorisine dâhil edilmiştir.

Makine öğrenmesi analizleri için iki farklı veri seti hazırlanmıştır. Bu veri setleri tahmini sağlayan nitelikler açısından farklılık göstermektedir. “sus0”

olarak adlandırılan veri setinde veriyi anlama bölümünde açıklanan tüm tahmini sağlayacak nitelikler yer alırken; “sus1” biçiminde adlandırılan veri setinde tahmini sağlayacak nitelikler olarak yalnızca SUS-TR'deki 10 madde kullanılmıştır. Veri setinde yapılan bu ayrımla; SUS-TR'deki 10 maddeye ek kullanılan yaş, cinsiyet ve bölüm niteliklerinin analiz sonuçlarına etkisi incelenmiştir.



Şekil 1. “sus0” veri setlerinde hedef niteliğin sınıflarına ait frekans değerleri.



Şekil 2. “sus1” veri setlerinde hedef niteliğin sınıflarına ait frekans değerleri.

“sus0” veri setinde tekrar eden bir adet gözlem tespit edilmiş ve veri setinden çıkarılmıştır. Son durumda veri setinde hedef niteliğe ait sınıflar incelendiğinde, ÖBS'nin kullanılabilirliğini kabul edilemez olarak nitelendiren 204 (%63.16), kabul edilebilir olarak nitelendiren 119 (%36.84) öğrencinin olduğu tespit edilmiştir (Şekil 1). “sus1” veri setinde tekrar eden on üç adet gözlem tespit edilmiş ve veri setinden çıkarılmıştır. Son durumda veri setinde hedef niteliğe ait sınıflar incelendiğinde, ÖBS'nin kullanılabilirliğini kabul edilemez olarak nitelendiren 196 (%63.02), kabul edilebilir olarak nitelendiren 115 (%36.98) öğrencinin olduğu tespit edilmiştir (Şekil 2).

### 3.3 Model kurma

Bu çalışmada ele alınan problem genel hatlarıyla ÖBS kullanıcılarının ÖBS'nin kullanılabilirliği hakkındaki kararlarını “kabul edilebilir” ve “kabul

edilemez” gibi iki farklı kategoride sınıflandırmak olduğundan, çalışma kapsamında danışmanlı öğrenmeyi kullanan C4.5 Karar Ağacı Algoritması (C4.5 Decision Tree Algorithm), Basit Bayes Sınıflandırıcı (Naive Bayes Classifier) ve k-En Yakın Komşu Algoritması (k-Nearest Neighbor Algorithm - kNN) ile farklı modeller kurularak analizler gerçekleştirilmiştir. Bu algoritmalar aşağıda kısaca özetlenmiştir:

C4.5 Karar Ağacı Algoritması hem kategorik hem de sürekli değer alan nitelikleri sınıflandırma ağacına dâhil eden güçlü bir sınıflandırma aracıdır [25]. Ayrıca veri setinde eksik veri olması durumunda da karar ağacının nasıl oluşturulabileceğine ilişkin çözüm sunmaktadır [26]. Lim, Loh ve Shih’in [27] çalışmasına göre C4.5 algoritması; sınıflandırmada iyi doğruluk değerleri sağlarken, çalışmada kullanılan diğer veri madenciliği - makine öğrenmesi algoritmaları arasında en hızlı algoritma olarak gösterilmiştir [28].

Basit Bayes Sınıflandırıcı ise örneklerin tüm niteliklerinin verilen sınıf değerleri bağlamında birbirinden bağımsız olduğunu varsaymaktadır ve buna “Basit Bayes Varsayımı” (Naive Bayes Assumption) denilmektedir [29]. Bu algoritmanın veri madenciliği için en verimli ve etkili tümevarımsal öğrenme algoritmalarından biri olduğu iddia edilmektedir [30].

k-En Yakın Komşu Algoritması iyi bilinen sınıflandırıcılardan biri olup, sınıfı bilinmeyen bir örneğin sınıf değeri, eğitim veri setinde kendisine en yakın k örneğin en çok tekrar eden sınıf değeri olarak tespit edilmektedir [31].

“sus0” ve “sus1” veri setlerine, yukarıda özetlenen üç algoritma uygulanırken Hold-out performans değerlendirme yönteminden faydalanılmıştır. Veri setleri; %70’e %30, %80’e %20 ve %90’a %10 oranında üç farklı şekilde eğitim ve test veri seti olarak ayrılmıştır. Model performans değerlendirme ölçütlerinden; doğruluk (accuracy), hata oranı (error rate), duyarlılık (sensitivity), belirleyicilik (specificity), pozitif öngörü değeri - PÖD (positive predictive value), negatif öngörü değeri - NÖD (negative predictive value) ile F-ölçütü (F-measure) hesaplanarak model performansları birbiriyle kıyaslanmıştır. Kullanılan performans değerlendirme ölçütleri kısaca şu şekilde açıklanabilir [22,32]:

Algoritmanın tüm doğru tahminleri toplamının, tüm tahminlerin sayısına (test veri setindeki örnek

sayısına) bölünmesiyle doğruluk değeri, doğruluk değerinin 1’den çıkarılmasıyla da hata oranı elde edilir. Duyarlılık, doğru tahmin edilmiş pozitif örneklerin tüm pozitif örnek sayısına; belirleyicilik ise doğru tahmin edilmiş negatif örneklerin tüm negatif örnek sayısına oranıdır. PÖD, doğru tahmin edilen pozitif örneklerin tüm pozitif tahminlerin sayısına; NÖD ise doğru tahmin edilen negatif örneklerin tüm negatif tahminlerin sayısına oranı olarak tanımlanabilir. F-ölçüsü ise duyarlılık ve PÖD’ün harmonik ortalamasıdır.

Çalışma sonunda ÖBS’nin kullanılabilirliğinin tahmini için en iyi performansı gösteren makine öğrenmesi modeli öneri olarak sunulmuştur. Model performans değerlendirmesi ve en iyi modelin seçimine ilişkin sonuçlar “Bulgular ve Tartışma” bölümde verilmiştir. Analizler R programlama dili [33] ile RStudio’da [34] gerçekleştirilmiştir. Analizlerde caret [35], dprep [36], e1071 [37], RWeka [38,39] ve xlsx [40] R paketlerinden faydalanılmıştır.

#### 4 Bulgular ve tartışma

%70’e %30, %80’e %20 ve %90’a %10 oranları ile her iki veri setinde de yapılan analizler sonucunda ortalama Doğruluk, Hata ve F-ölçütü dikkate alınarak Tablo 1 oluşturulmuştur. Analizlerde, doğruluk ve hata değerleri birbirine çok yakın olsa da, duyarlılık ve pozitif öngörü değerinin kullanılarak hesaplanan F-ölçütü değeri 80/20 ayrımını diğer analizlerin bir adım önüne geçirmiştir. Tablo 1’e göre veri seti gözetmeksizin yapılan değerlendirme sonucunda en iyi ortalama performansın %80’e %20 ayrımıyla elde edildiği tespit edilmiş (Doğruluk = 0.590, F-ölçütü = 0.700) ve bu bölümde 80/20 Hold-out performans değerlendirme yöntemine ait performans değerlendirme sonuçları paylaşılmıştır.

Tablo 1. Farklı Hold-out oranları için elde edilen ortalama performans değerleri.

Hold-out Oranı	Ortalama		
	Doğruluk	Hata	F-ölçütü
70/30	0.581	0.419	0.696
80/20	0.590	0.410	0.700
90/10	0.591	0.409	0.694

##### 4.1 C4.5 karar ağacı algoritmasından elde edilen bulgular

C4.5 Karar Ağacı Algoritmasından elde edilen performans değerlendirme ölçülerine göre (Tablo 2) algoritma her iki analizde de test veri setindeki tüm örnekleri frekans değeri daha yüksek olan

“kabul edilemez” sınıfına ait olarak bulmuştur. “kabul edilebilir” sınıfından hiçbir örnek tahmin edilemediği için negatif öngörü değerleri de tanımsız çıkmıştır. Algoritmanın performansı tatmin edici olmadığından modelden elde edilen karar ağacı ve kurallara çalışmada yer verilmemiştir.

#### 4.2 Basit Bayes sınıflandırıcıdan elde edilen bulgular

Basit Bayes sınıflandırıcıdan elde edilen bulgulara göre (Tablo 3) en yüksek performans değerleri “sus0” veri setinden elde edilmiştir. Kullanılabilirlik için “kabul edilemez” sınıfına ilişkin sonuçlar “kabul edilebilir” sınıfına göre daha iyi olduğu gözlemlenmiştir.

#### 4.3 k-En yakın komşu algoritmasından elde edilen bulgular

k-En Yakın Komşu Algoritması için komşu sayısını belirten k değeri 1’den 20’ye kadar denenmiştir. Veri setleri için doğruluk ve F-ölçütünün k parametresinin farklı değerlerine göre nasıl değiştiği sırasıyla Şekil 3’te ve Şekil 4’te verilmiştir. “sus0” ve “sus1” veri setleri için en iyi performansın elde edildiği analiz sonuçları grafik üzerinde belirtilmiştir. “sus0”dan elde edilen performans ölçülerinin “sus1”den önde olduğu görülmektedir.

Analizlerde k’nın 1 ile 20 arasında aldığı değerler elde edilen doğruluk ve F-ölçütü değerleri açısından belirgin olumlu bir fark yaratmadığını

Tablo 2. C4.5 Karar Ağacı Algoritmasından elde edilen performans değerleri.

Veri Seti	Doğruluk	Hata	Duyarlılık	Belirleyicilik	PÖD	NÖD	F-ölçütü
sus0	0.635	0.365	1	0	0.635	-	0.777
sus1	0.629	0.371	1	0	0.629	-	0.772

Tablo 3. Basit Bayes Sınıflandırıcıdan elde edilen performans değerleri.

Veri Seti	Doğruluk	Hata	Duyarlılık	Belirleyicilik	PÖD	NÖD	F-ölçütü
sus0	0.698	0.302	0.850	0.435	0.723	0.625	0.782
sus1	0.645	0.355	0.795	0.391	0.689	0.529	0.738

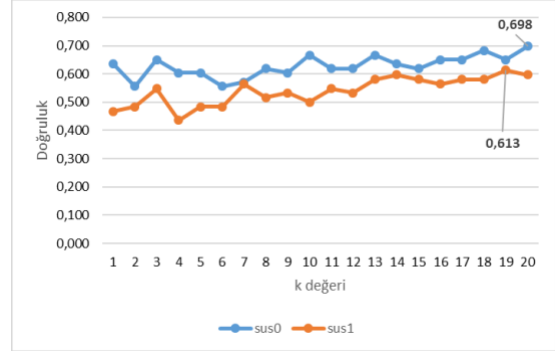
Tablo 4. k-En Yakın Komşu Algoritmasından elde edilen performans değerleri

Veri Seti	k	Doğruluk	Hata	Duyarlılık	Belirleyicilik	PÖD	NÖD	F-ölçütü
sus0	20	0.698	0.302	0.925	0.304	0.698	0.700	0.796
sus1	19	0.613	0.387	0.949	0.043	0.627	0.333	0.755

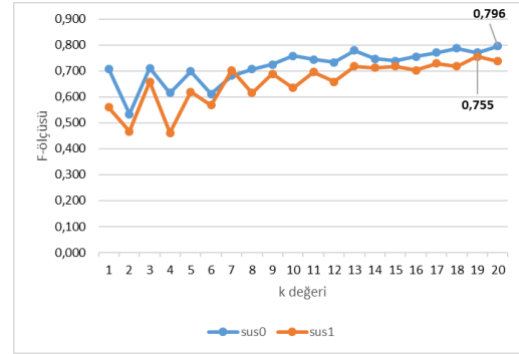
## 5 Sonuçlar

Bu çalışmada, sistem kullanılabilirliği kararının tahmin edilebilmesi için yapay zekânın alt çalışma alanlarından biri olan makine öğrenmesi kullanılmıştır. Bu kapsamda, Türkiye’deki bir vakıf

göstermektedir; ancak Tablo 4’te her iki veri seti için elde edilen en iyi performans değerleri daha detaylı bir biçimde verilmiştir. Analizlerde sistemin kullanılabilirliğine ilişkin karar sınıflarına özgü elde edilen performans değerleri arasındaki fark Tablo 4’ten de görülebilmektedir.



Şekil 3. k-En Yakın Komşu Algoritması analizlerinden elde edilen doğruluk değerleri



Şekil 4. k-En Yakın Komşu Algoritması analizlerinden elde edilen F-ölçütü değerleri

üniversitesine ait ÖBS’nin kullanılabilirliği üzerine inceleme yapılmıştır. Bu inceleme için öğrencilerin yaş, cinsiyet ve bölüm ve SUS-TR ölçeğine verdikleri yanıtlar kullanılarak “sus0”, sadece SUS-TR ölçeğine verdikleri yanıtlar kullanılarak da

“sus1” adında iki veri seti oluşturulmuştur. Oluşturulan veri setleri üzerinde C4.5 Karar Ağacı Algoritması, Basit Bayes Sınıflandırıcı ve k-En Yakın Komşu algoritması ile analizler gerçekleştirilmiştir. Hold-out yöntemiyle %70’e %30, %80’e %20 ve %90’a %10 oranları ile yapılan analizlerde en iyi ortalama performans 80/20 ayırımından elde edildiği için (Doğruluk = 0.590, F-ölçütü = 0.700), çalışmadaki tüm detaylı performans değerlendirmeleri 80/20 ayırımı için paylaşılmıştır.

Elde edilen bulgulara göre, öğrencilerin yaş, cinsiyet ve bölüm ve SUS-TR ölçeğine verdikleri yanıtlar kullanılarak oluşturulan “sus0” veri setinde algoritmaların genel olarak daha iyi performans gösterdiği tespit edilmiştir; ancak yapılan bu inceleme SUS-TR’ye verilen yanıtlar haricinde yalnızca üç farklı niteliği içermektedir. İleriki çalışmalarda nitelik sayısının artırılarak daha kapsamlı bir veri seti ile çalışılabilir. Ayrıca, öğrencilerin sistemin kullanılabilirliği üzerine kararlarına etkisi olduğu düşünülen niteliklerin etki dereceleri incelenebilir, nitelik seçimi yöntemlerinden faydalanılarak en etkili niteliklerle performans değerlendirmesi yapılabilir. Yaş, cinsiyet ve öğrencilerin bilişimle ilgili bir bölümde öğrenim durumu niteliklerinin modelleri daha açıklayıcı hale getirdiğine inanılmaktadır; ancak bu durumun nedeninin incelenmesi bir başka çalışmanın araştırma konusu kapsamına girmektedir. Sisteme ait kullanılabilirliğin belirlenebilmesi için literatürde System Usability Scale (SUS) [23], Questionnaire for User Interface Satisfaction (QUIS) [41] ve Computer System Usability Questionnaire (CSUQ) [42] gibi belirli başlı ölçekler olsa da, bir sistemin kullanılabilirliğinin öznel bir yanı da bulunmaktadır. Bu nedenle, sistem kullanılabilirliği analizleri için ne kadar o sistemin kullanıcılarına (ya da hedef kitlesine) özgü nitelik toplanarak analizlerde göz önünde bulundurulursa, o kadar iyi analiz sonuçlarının elde edileceğine inanılmaktadır.

Bulgular incelendiğinde, doğruluk dışındaki diğer performans değerlendirme ölçüleri de göz önünde bulundurulduğunda her iki veri seti için de en iyi sonuç veren algoritmanın k-En Yakın Komşu Algoritması olduğu görülmüştür (k=20 için doğruluk = 0.698, F-ölçütü = 0.796). Bu noktada, algoritmanın tembel öğrenmeden (*lazy learning*) dolayı sabit bir model üretmeden her seferinde eğitim veri setindeki örneklere bağlı kalarak

hesaplamaları tekrar tekrar yapması, ileride daha fazla nitelik ve gözlemlerle gerçekleştirilecek çalışmalar için bir dezavantaj olarak görülmektedir. “sus0” veri seti için k’nın 1’den 20’ye kadar olan değerleriyle elde edilen doğruluk ve F-ölçütü değerleri açısından belirgin olumlu bir fark yaratmadığı, algoritma performansının 0.5 ile 0.7 bandında değiştiği gözlemlenmiştir. Yapılan analizlerde en düşük performansı C4.5 Karar Ağacı Algoritması göstermiştir. Frekans değeri daha düşük olan ve ÖBS’nin kullanılabilirlik durumunu “kabul edilebilir” olarak belirten sınıfı diğerinden ayırmakta başarısız olduğu görülmüştür. İleride gerçekleştirilecek çalışmalarda yapay sinir ağları, destek vektör makinaları gibi daha farklı ve ileri düzey olarak görülen makine öğrenmesi algoritmalarıyla çalışılarak daha iyi sonuçlar elde edilebilir.

İnsan bilgisayar etkileşimi çalışmalarında faydalanılan göz izleme cihazlarının ürettiği veri makine öğrenmesi teknikleri kullanılarak sistem kullanılabilirliği analiz edilebilir. Makine öğrenmesi teknikleriyle gerçekleştirilen analizler sonucunda elde edilen en iyi performansa sahip model bir kullanıcı arayüzü ile sistem kullanıcılarıyla paylaşılarak, sistemin kullanılabilirlik testleri için alternatif bir yöntem sunulabilir. Böylelikle, CRISP-DM’nin son aşaması olan modelin uygulamaya geçirilmesi adımı da tamamlanmış olacaktır.

### Kaynaklar

- [1] ISO 9241-11. “Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability”. International Organization for Standardization, 1998.
- [2] Jordan PW. "An Introduction to Usability". Padstow, UK, Taylor & Francis Ltd., 1998.
- [3] Nielsen J. “Usability Engineering”. San Diego, CA, USA, Morgan Kaufman Academic Press, 1993.
- [4] Nielsen J. “Usability 101: Introduction to Usability”, 2012. <https://www.nngroup.com/articles/usability-101-introduction-to-usability/> adresinden 3 Ekim 2018 tarihinde erişildi.
- [5] Hornbæk K. “Current practice in measuring usability: Challenges to usability studies and research”. International Journal of Human-Computer Studies, 64(2), 79–102, 2006.
- [6] Sweeney M, Maguire M, Shackel B. “Evaluating user-computer interaction: a framework”. Int J Man Mach Stud, 38(4), 689–711, 1993.
- [7] Nielsen J. "Finding usability problems through heuristic evaluation". Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, ACM, 373–80, 1992.



- [8] Rubin J, Chisnell D. "Handbook of usability testing: how to plan, design, and conduct effective tests". 2nd ed. Indianapolis, USA, Wiley Publishing Inc., 2008.
- [9] Kieras D. "Model Based Evaluation". In *The Human-Computer Interaction Handbook Fundamentals, Evolving Technologies, and Emerging Applications*, 2nd ed., Sears A, Jacko JA. Eds. Boca Raton, USA, CRC Press, 2007.
- [10] Mukerjee S. "Student information systems - implementation challenges and the road ahead". *Journal of Higher Education Policy and Management*, 34(1), 51-60, 2012.
- [11] Carcary M, Long G, Remenyi D. "The Implementation of a New Student Management Information System (MIS) at an Irish Institute of Technology -- An Ex Post Evaluation of its Success". *Electronic Journal of Information Systems Evaluation*, 10(1), 14, 2007.
- [12] Karat CM. "A Business Case Approach to Usability Cost Justification for the Web". In *Cost-justifying usability: an update for an Internet age*, 2nd ed., Bias RG, Mayhew DJ. Eds. San Francisco, USA, Morgan Kaufman Publishers, 2005.
- [13] Brooke J. "SUS - A quick and dirty usability scale". *Digital Equipment Corporation*, 7, 1986.
- [14] Demirkol D. "Evaluation of student information system (SIS) in terms of user emotions, performance, and perceived usability: A Turkish university case". Yüksek Lisans Tezi, Yeditepe Üniversitesi, İstanbul, 2018.
- [15] Korvald C, Kim E, Reza H. "Evaluation and implementation of machine learning techniques in usability testing for web sites". In *Proceedings of the 47th Annual Midwest Instruction and Computing Symposium*, Verona, WI, 1-11, 2014.
- [16] Oztekin A. "A decision support system for usability evaluation of web-based information systems". *Expert Syst Appl*, 38(3), 2110-2118, 2011.
- [17] Oztekin A, Delen D, Turkyilmaz A, Zaim S. "A machine learning-based usability evaluation method for eLearning systems". *Decis Support Syst*, 56, 63-73, 2013.
- [18] Davis PA. "Learning usability assessment models for web sites". Doctoral Thesis, Texas A&M University, USA, 2010.
- [19] Horng D, Kittur A, Hong JI, Faloutsos C. "Making Sense of Large Network Data: Combining Rich User Interaction and Machine Learning". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, BC, Canada, 167-176, 2011.
- [20] Maragoudakis M, Tselios NK, Fakotakis N, Avouris NM. "Improving SMS Usability Using Bayesian Networks". In *Methods and Applications of Artificial Intelligence*, 2308, Vlahavas IP, Spyropoulos CD. Eds. Berlin, Springer, 179-190, 2002.
- [21] Shearer C. "The CRISP-DM model: the new blueprint for data mining". *Journal of data warehousing*, 5(4), 13-22, 2000.
- [22] Balaban ME, Kartal E. *Veri Madenciliği ve Makine Öğrenmesi Temel Algoritmaları ve R Dili ile Uygulamaları*. 2nd ed. Beyoğlu, İstanbul, Çağlayan Kitabevi, 2018.
- [23] Brooke J. "SUS - A quick and dirty usability scale". In *Usability Evaluation In Industry*, Jordan PW, Thomas B, McClelland IL, Weerdmeester B. Eds. London, UK, Taylor & Francis Ltd., 189-194, 1996.
- [24] Bangor A, Kortum P, Miller J. "Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale". *Journal of Usability Studies*, 4(3), 114-123, 2009.
- [25] Selçukcan Erol Ç. "Sağlık Bilimlerinde R ile Veri Madenciliği". In *R ile Veri Madenciliği Uygulamaları*. 1st ed., Balaban ME, Kartal E. Eds. İstanbul, Çağlayan Kitabevi, 2016.
- [26] Quinlan JR, "Constructing Decision Trees". In *C4.5 Programs for Machine Learning*, California, USA, Morgan Kaufman Publishers, 17-26, 1993.
- [27] Lim TS, Loh WY, Shih YS. "A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms". *Machine Learning*, 40(3), 203-228, 2000.
- [28] Ruggieri S. "Efficient C4.5". *IEEE Transactions on Knowledge and Data Engineering*, 14(2), 438-444, 2002.
- [29] McCallum A, Nigam K. "A Comparison of Event Models for Naive Bayes Text Classification". *AAAI-98 Workshop on Learning for Text Categorization*, 752, 41-48, 1998.
- [30] Zhang H. "The optimality of naive Bayes". In *Proceedings of 17th International Florida Artificial Intelligence Research Society Conference*, Florida, USA, 562-567, 2004.
- [31] Gou J, Xiong T, Kuang Y. "A Novel Weighted Voting for K-Nearest Neighbor Rule". *Journal of Computers*, 6(5), 833-840, 2011.
- [32] Han J, Kamber M, Pei J. "Data Mining: Concepts and Techniques". 3rd ed. MA, USA: Morgan Kaufman Publishers, 2012.
- [33] The R Foundation. "R: The R Project for Statistical Computing". 2018. <https://www.r-project.org/adresinden> 14 Kasım 2018 tarihinde erişildi.
- [34] RStudio. "RStudio - Open source and enterprise-ready professional software for R". RStudio, 2014. <https://www.rstudio.com/> adresinden 14 Kasım 2018 tarihinde erişildi.
- [35] Kuhn M. "caret: Classification and Regression Training". 2018.
- [36] Acuna E, The CASTLE research group. "dprep: Data Pre-Processing and Visualization Functions for Classification". 2015.
- [37] Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. "e1071: Misc Functions of the Department of Statistics", Probability Theory Group (Formerly: E1071), TU Wien. 2017.
- [38] Hornik K, Buchta C, Zeileis A. "Open-Source Machine Learning: R Meets Weka". *Computational Statistics*, 24(2), 225-232, 2009.
- [39] Witten IH, Frank E. "Data Mining: Practical Machine Learning Tools and Techniques", 2nd ed. San Francisco, USA, Morgan Kaufmann, 2005.
- [40] Dragulescu AA, Arendt C. "xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files". 2018.
- [41] Harper BD, Norman KL. "Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5". In *Proceedings of the 1st Annual Mid-Atlantic Human Factors Conference*, 224-228, 1993.
- [42] Lewis JR. "IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use". *Int J Hum-Comput Int*, 7(1), 57-78, 1995.