Süleyman Demirel Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
Cilt 22, Sayı 3, 1249-1255, 2018

Süleyman Demirel University
Journal of Natural andAppliedSciences
Volume 22, Issue 3, 1249-1255, 2018

DOI: 10.19113/sdufenbed.501503

Tuş Vuruş Dinamikleri ile Klavye Kullanım Stiline Dayalı Kullanıcı Tanıma

İslam Mayda*[1], İbrahim Demir[2]

[1]Yıldız Teknik Üniversitesi, Elektrik-ElektronikFakültesi, Bilgisayar Mühendisliği Bölümü, 34220, İstanbul
[2]Yıldız Teknik Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, 34220, İstanbul

(Alınış / Received: 29.03.2018, Kabul / Accepted: 05.12.2018, Online Yayınlanma/ Published Online: 24.12.2018)

Anahtar Kelimeler
Kullanıcı tanıma,
Klavye kullanım stili,
Tuş vuruş dinamikleri,
Biyometri,
Makine öğrenmesi

Özet: İnternet siteleri ve internet tabanlı sistemlerde kullanıcılara özel olarak açılan hesaplara, yine kullanıcılara özel şifrelerle giriş sağlanmaktadır. Ancak, bu şifreler çeşitli şekillerde başkaları tarafından ele geçirilebilmekte ve hesabın gerçek sahibi dışındaki kişiler de bu hesaplara girebilmektedir. Bu durumu önlemek üzere güvenliğin bir kat daha arttırılması için bu çalışmada, şifreyle birlikte kullanıcıların klavye kullanım stilinin de kullanılabileceği önerilmektedir. Çalışmada, bir denek grubu ile yapılan testlerin sonuçları sunularak, değerlendirmeler paylaşılmıştır. Yüksek bir başarı oranı ile klavye kullanım stiline dayalı olarak kullanıcıları tanımanın mümkün olduğu gösterilmiştir. Kullanıcı tanımada sınıflandırma başarısını arttırmak için birden fazla verinin birleştirilerek tek bir veri olarak kullanılmasının mümkün olduğu ortaya konmuştur. Birleştirilmiş verilerin kullanıldığı sistemlerde yeni bir kullanıcı için sisteme çok daha az sayıda veri girişi yaptırarak sistemin yeni kullanıcıyı öğrenmesinin sağlanabileceği gösterilmiştir.

User Identification Based on Keyboard Usage Style with Keystroke Dynamics

Keywords
User identification,
Keyboard usage style,
Keystroke dynamics,
Biometrics,
Machine learning

Abstract: Access to the accounts which are created user specifically on internet sites and internet based systems, is provided again with user-specific passwords. However, these passwords can be captured by others in various ways and persons other than the real owner of the account can also enter these accounts. In order to increase the security one more time to prevent this situation, in this work, it is suggested that the keyboard usage style of the users can also be used with the password. In the study, evaluations were shared by presenting the results of tests made with a group of subjects. It has been shown that it is possible to recognize users based on the keyboard usage style with a high success rate. It is proved that it is possible to use more than one data as a single data by merging them to increase the classification efficiency in user identification. It has been presented that it is possible for the system to learn a new user by making a much smaller number of data entries for the user in the system which uses combined data.

1. Giriş

Sadece belirli bir kişiye veya gruba ait olan verilerin başkaları tarafından erişilememesi için çeşitli kullanıcı giriş sistemleri kullanılmaktadır. Kullanıcı giriş sistemlerinde yaygın olarak bir kullanıcı adı ile şifre veya e-posta adresi ile şifre ikilisi tercih edilmektedir. Genelde kullanıcı adı ve e-posta adresleri başka kişiler tarafından da görülebildiği için, kullanıcıya özel olarak sadece şifre bilgisi kalmaktadır.

Kullanıcılara ait olan bu şifre çeşitli şekillerde başka birilerinin de eline geçebilmektedir. Örneğin, bir kişi şifresini yazarken başka biri fark ettirmeden onu izleyebilir veya kullanıcıdan izinsiz olarak bilgisayarına kurulan tuş kaydedici (key logger) gibi casus yazılımlarla şifre başkalarının eline geçebilir. Özellikle 12345, 123456, 1234567, 12345678 gibi yaygın olarak tercih edilen [1] tahmin edilebilir ve kırılması kolay bir şifre seçen kullanıcıların verilerinin çalınması daha muhtemeldir. Bireylerin şifre yapılarına yönelik yapılan bir araştırmada [2] şifrelerini yalnızca rakamlardan oluşturan kullanıcıların çoğunluğunun şifre yapılarında telefon numarası, kimlik numarası, doğum tarihi, okul

*İlgili yazar: islam.mayda@stu.khas.edu.tr

numarası gibi yine başkaları tarafından bilinebilen bilgileri kullandıkları görülmüştür.

Kullandığımız şifrelerin bir şekilde başkalarının eline geçmesi demek e-postalarımızın, banka hesaplarımızın, kişisel bilgisayarımızın, sosyal medya hesaplarımızın, WiFi ağımızın, e-Devlet Kapısı gibi şahsımıza ait çok özel bilgilerin yer aldığı kamu kurumu hesaplarımız veya üniversite bilgi yönetim sistemi hesabımız gibi şifre ile giriş sağladığımız tüm sistemlerde bize ait olan her şeyin başkasının eline geçmesi anlamına gelmektedir. Özellikle askeri yazılımlar gibi kritik bilgilerin saklandığı sistemlere yabancıların girmesi hayati sonuçlar doğurabilir. Asıl kullanıcı kendi şifresini bilerek başka biriyle paylaşmış da olabilir ve bu durumda bile asıl kullanıcı dışındaki yabancı birinin sisteme girmesi istenmeyebilir. Kullanıcı giriş sistemleri bizim için bu kadar önemliyken sadece tek bir şifre ile güvenliği sağlamak çoğu zaman yeterli olmayabilir. Bu yüzden, bu çalışmada kullanıcıları tanımada şifrelerin yanı sıra ek bir güvenlik sistemi önerilmektedir.

Klavye kullanım stilinin biyometrik özelliğine dair son yıllarda onlarca çalışma yapılmıştır. Bu çalışmalarda temel olarak, kullanıcıların tuş vuruş dinamiklerinin onları tanımada ve ayırt etmede ne kadar başarılı olduğu incelenmiştir. Bu bölümün devamında bu alanda daha önce yapılmış olan çalışmalara dair bir özet verilirken, Materyal ve Metot bölümünde bu çalışmada kullanılan yöntem anlatılmış, toplanan veriler üzerinde yapılan testler Bulgular başlığı altında sunulmuş, Tartışma ve Sonuç bölümünde de çalışma sonucuna dair değerlendirmeler paylaşılmıştır.

Tuş vuruş doğrulama tekniği statik veya sürekli olarak sınıflandırılmaktadır. Statik doğrulama tekniğinde doğrulama işlemi kullanıcı giriş sistemleri gibi sadece belirli bir aşamada yapılmaktadır. Statik yaklaşım sıradan şifrelere göre daha sağlam bir doğrulama sunsa da sürekli güvenlik sağlamamaktadır ve girişten sonra klavye başındaki kullanıcının değişmesi durumunda bunu fark edememektedir. Sürekli doğrulama tekniğinde ise, etkileşim devam ettiği sürece kullanıcının klavye kullanımını izlemektedir [3]. Eğer kullanıcı sisteme giriş yaptıktan sonra klavyeyi çok kullanmadan, genelde veri görüntüleme yapıyorsa burada statik yaklaşım uygulanabilir. Örneğin, kullanıcı banka hesabını kontrol etmek için internet sitesi üzerinden giriş yaptıktan sonra metin girişi yapmasına pek ihtiyaç olmamaktadır [4]. Diğer taraftan, sosyal medya hesabı veya mesajlaşma uygulamaları gibi kullanıcının daha çok metin girişi yaptığı sistemlerde dinamik yaklaşım uygulanması mümkündür. Dinamik doğrulama tekniğinde kullanıcının bütün etkileşimi takip edildiğinden kullanıcıyı tanımanın daha kolay olması beklenir. Diğer yandan, statik doğrulama tekniğinde ise çok daha kısıtlı bir veri kullanıcıyı tanımada kullanılmaktadır. Bu çalışmada statik doğrulama üzerinde çalışılması tercih edilmiştir ve bu yaklaşıma dair yapılan daha önceki çalışmalar bu bölümde sunulmaktadır.

Tuş vuruş doğrulama tekniğinin performansı genel olarak hata oranlarına göre değerlendirilmektedir. Burada kullanılmakta olan iki farklı hata oranı şunlardır: Yanlış Kabul Oranı (False Acceptance Rate - FAR) ve Yanlış Ret Oranı (False Rejection Rate - FRR). FAR yanlış bir kişinin doğru kişi sanılarak güvenli bir sisteme başarıyla giriş yapabilme ihtimalini ifade etmektir. İstatistikte buna Tip II hatası da denilmektedir. FRR ise doğru kişilerin yanlış kişi sanılarak sisteme girişinin engellenme yüzdesidir. İstatistikte bu değer Tip I hatası olarak da tanımlanmaktadır. İdeal durumda iki hata oranı da %0 olmalıdır [5].

Tuş vuruş dinamikleri kullanılarak yapılan araştırmalar literatür tarama çalışmalarında özetlenmiştir [4, 5, 6]. Bu çalışmalarda görüldüğü gibi istatistiksel ve yapay sinir ağlarına dayalı sınıflandırma yöntemleri yoğun olarak kullanılmıştır. Yapay sinir ağlarına dayalı sınıflandırma yöntemlerinin istatistiksel sınıflandırma yöntemlerine göre genelde daha yüksek başarı gösterdiği görülmektedir. Tuş vuruş zaman verisi olarak da genelde; bir tuşun basılı kalma süresi, bir tuşun bırakılması ile bir sonraki tuşun basılması arasında geçen süre, bir tuşun basılması ile bir sonraki tuşun basılması arasında geçen süre, bir tuşun bırakılması ile bir sonraki tuşun bırakılması arasında geçen süre, bir tuşun basılması ile iki sonraki tuşun basılması arasında geçen süre gibi metrikler kullanılmıştır.

Hangi metriğin sınıflandırmada daha başarılı olduğunu görmek için bazı çalışmalar yapılmıştır [7, 8]. Bu çalışmalarda; bir tuşun basılı kalma süresi ile iki tuş arasındaki süre metrikleri hem tek başlarına hem de bir arada kullanılarak sınıflandırma yapılmış ve en başarılı sonuçların iki metrik birlikte kullanıldığında alındığı görülmüştür.

Tuş vuruş dinamikleri sadece kullanıcı tanımda değil, farklı amaçlar için kullanılması da araştırılmıştır. Örneğin, Fairhurst ve Costa-Abreu tarafından yapılan çalışmada [9] sosyal ağlarda kullanıcıların cinsiyetinin tahmin edilmesinde klavye kullanım stilinin kullanılabileceğini gösterilmiştir. Ayrıca, bu çalışmada aynı yöntemle cinsiyetin yanı sıra yaş aralığı gibi kullanıcıya dair farklı özelliklerin de tahmin edilebileceği vurgulanmaktadır.

2. Materyal ve Metot

2.1. Şifrenin oluşturulması

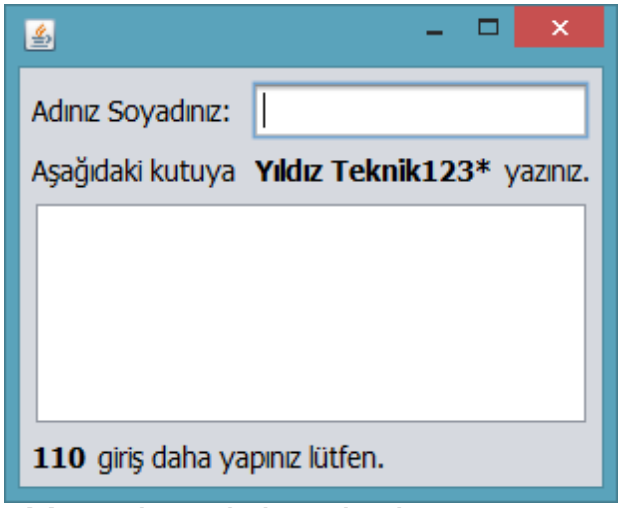
Güçlü bir şifre oluşturulması için Google, şifre içerisinde büyük/küçük harf, rakam ve simgelerin karışımının kullanılmasını önermektedir [10].

Microsoft da güçlü bir parolanın en az sekiz karakter uzunluğunda olması ve büyük harf, küçük harf, sayı ve semboller içermesi gerektiğini belirtmektedir [11]. Birçok web sitesi kullanıcılarını bu önerilerde belirtilen özelliklere sahip parola oluşturmaya zorlamaktadır. Bu önerilerden yola çıkarak çalışmada kullanılmak üzere içinde büyük harf, küçük harf, rakam, boşluk ve sembol içeren sabit metin olarak *Yıldız Teknik123** belirlenmiştir.

2.2. Veri toplama

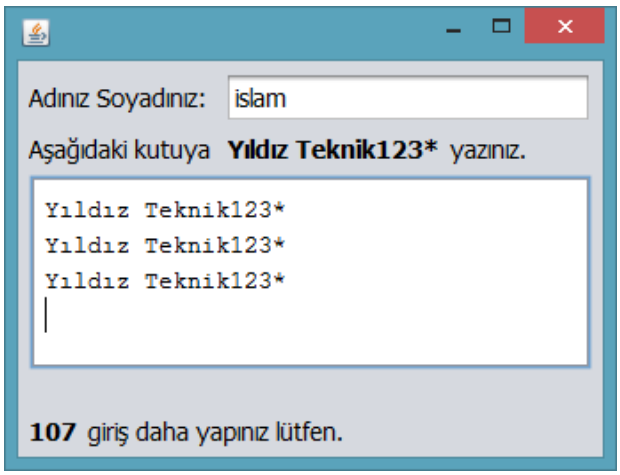
Çalışmada kullanılacak verileri toplamak için 10 denek seçilmiş ve kendilerinden, belirlenen sabit metni 110 kez yazmaları istenmiştir. Denek sayısına karar verilirken daha önce yapılan çalışmalarda kullanılan denek sayıları dikkate alınmıştır. [4]'te derlenen, örüntü tanıma ve öğrenme tabanlı algoritmaların kullanıldığı kimlik doğrulama ve tanıma çalışmalarının sunulduğu tablodaki denek sayılarına bakıldığında bu sayının en az 8, en fazla da 315 olduğu görülmektedir. Derleme çalışmada sunulan bu tabloya göre denek sayısı belirtilmiş olan 37 çalışmadan 8 tanesinde bu sayı 10 veya daha azdır. Buna bağlı olarak, veri toplama konusundaki zorluklar da göz önünde bulundurulmuş ve denek sayısı 10 olarak belirlenmiştir.

Deneklerden veri toplama işlemi için Java programlama dilinde bir uygulama yazılmıştır. Uygulamanın başlangıçtaki açılış arayüzü Şekil 1'deki gibidir.



Şekil 1. Uygulamanın başlangıçtaki açılış arayüzü

Uygulama ilk açıldığında denekten üstteki metin kutusuna kendi ismini yazması istenir. Daha sonra alttaki metin alanına *Yıldız Teknik123** metnini yazar ve ENTER tuşuna basarak bir alt satıra geçer, tekrar aynı metni yazar ve tekrar ENTER tuşuna basar. 110 giriş yaptıktan sonra uygulama kapatılır ve bir denekten veri alma işlemi tamamlanmış olur. Bir deneğin veri girişine ait görsel Şekil 2'de sunulmuştur.



Şekil 2. Veri girişine ait bir örnek

Şekil 2'de görüldüğü gibi denek belirlenen şifreyi üç kez girmiş ve kendisinden 107 kez daha giriş yapması beklenmektedir.

Tüm veri standart Q klavye dizilimine sahip bir dizüstü bilgisayarla toplanmıştır. Farklı klavyelerde tuşların arasındaki mesafeler değişebildiğinden dolayı veri toplamada birden fazla bilgisayar veya klavye kullanılmamıştır.

Bilgisayar kullanıcıları kendi şifrelerini gündelik hayatlarında sık sık kullandıkları için şifre girerken çok fazla hata yapmazlar. Bu varsayımdan yola çıkarak deneklerin veri girişi sırasında yaptığı hatalı girişler, denek BACKSPACE tuşu ile hatalı girdiyi silinerek veri girişini düzgün şekilde tamamlansa dahi, kabul edilmemiştir. Zira hatalı girişin silinmesi ile doğru tuşa basılması arasında geçen süre o giriş sırasında alınan verinin doğallığını bozmaktadır.

2.3. Denekler

Çalışmada kendisinden veri alınan denekler, günlük hayatlarında hemen her gün bilgisayar kullanan kişilerdir. Deneklere dair bazı bilgiler Tablo 1'de listelenmiştir.

Tablo 1. Deneklerin özellikleri

Denek Adı	Yaşı	Mesleği	Büyük harf için tercihi
İbrahim	46	Akademisyen	SHIFT
İslam	27	Bilgisayar Mühendisi	CAPS LOCK
İsmail	33	Dijital Pazarlamacı	SHIFT
Kemal	21	Geliştirici	SHIFT
Kerem	27	Elektronik Mühendisi	CAPS LOCK
Mirsat	27	Akademisyen	SHIFT
Ömer	26	Muhasebeci	SHIFT
Samet	23	Bilgisayar Mühendisi	CAPS LOCK
Serdar	27	İş Analisti	SHIFT
Yunus	30	3D Modelci	CAPS LOCK

Deneklerin metinde geçen büyük harfleri yazarken ya CAPS LOCK tuşunu ya da SHIFT tuşunu kullanmaktadır. Bir deneğin veri girişi sırasında ikisinden birini tercih ettiği gözlenmiş, bazı verilerde birinci yöntemi bazılarında ise diğer yöntemi kullanan bir deneğe rastlanmamıştır.

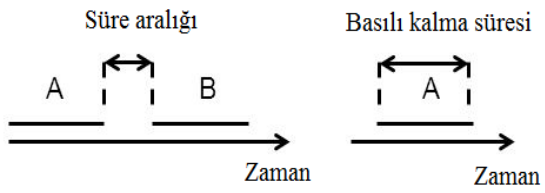
Büyük harf yazarken yapılan tercihin haricinde, rakamları ve yıldız sembolümü yazarken de bir tercih yapılmıştır. Deneklerin bir kısmı klavyede fonksiyon tuşları ile harf tuşları arasında tek sıra halinde yatay sıralanan rakam tuşları ve yıldız tuşunu kullanırken, diğer bir kısmı da klavyenin sağ tarafında kalan yoğun matematiksel işlemler için ayrılmış dört sütundan oluşan sayısal alanı kullandığı görülmüştür.

2.4. Verinin yapısı

Denekler belirlenen metni yazarken klavye kullanım stilini öğrenmek üzere aşağıdaki iki özelliğe dikkat edilmiştir:

1. Bir tuşun basılı kalma süresi
2. Bir tuştan diğer tuşa geçiş süresi

Birinci özellik kullanıcının bir tuşa basması ile parmağını o tuşun üzerinden kaldırması arasındaki süreyi ifade ederken, ikinci özellik kullanıcının parmağını bir tuşun üzerinden kaldırması ile bir sonraki tuşa basması arasındaki süreyi ifade etmektedir. Bu süreler milisaniye cinsinden hesaplanarak kaydedilmiştir. Bu özellikler Şekil 3'te görsel olarak ifade edilmiştir.



Şekil 3. Klavye kullanım stili özellikleri

Denek 17 karakter içeren metni girerken büyük harf yapmak için CAPS LOCK tuşunu kullanırsa, metin iki adet büyük harf içerdiğinden dolayı toplam 21 tuş basımı yapar ve son olarak ENTER tuşuna basar. Her tuş için tuşun basılı kalma süresi ve tuşun bırakılması ile bir sonraki tuşun basılması arasındaki süre olmak üzere ikişer özellik olmak üzere 42 özellik ve son olarak ENTER tuşunun basılı kalma süresi ile birlikte toplam 43 özellik kaydedilmiştir. Bir deneğe ait üç adet girdi verisi Şekil 3'te sunulmuştur.

Eğer denek metni girerken büyük harfleri SHIFT tuşunu kullanarak yazarsa toplam 39 özelliği olan veri elde edilmektedir. CAPS LOCK ve SHIFT tuşunu kullanan deneklerin verilerini birlikte değerlendirebilmek için SHIFT tuşu kullanılan verilerin sonuna "0" şeklinde 4 adet boş veri eklenmiştir.

```
46 @data
47 172,31,94,0,141,31,78,110,125,15,78,32,93
   ,0,141,31,125,110,125,15,63,31,47,148,109
   ,-15,109,110,78,-16,156,-31,94,156,141,93
   ,79,93,110,0,78,94,109,islam
48 171,88,61,16,141,31,78,125,109,-47,94,47,
   94,-16,141,15,141,250,203,375,78,-15,108,
   93,141,-47,63,140,47,94,141,-63,125,31,12
   5,86,94,94,140,-47,79,78,125,islam
49 172,110,94,-32,172,-15,78,125,93,0,79,62,
   78,-31,125,47,94,31,94,-1,63,42,91,125,11
   0,-31,78,125,62,78,125,-47,125,32,125,78,
   94,109,78,-31,63,77,96,islam
```

Şekil 4. Bir deneğe ait üç adet girdi verisi

Şekil 4'te görülen girdi verisindeki bazı değerlerin eksi olması, basılı olan bir tuş bırakılmadan sonraki tuşa basılması sebebiyledir.

2.5. Verinin analizi

Toplanan veriler altı farklı makine öğrenmesi yöntemi ile sınıflandırılmış ve yöntemlerin sınıflandırma başarıları değerlendirilmiştir. Kullanılan yöntemler şunlardır: Naive Bayes, Çok Katmanlı Algılayıcı, DVM (Destek Vektör Makinesi), Bagging, Karar Ağacı, Rassal Orman. Sınıflandırma işlemi için veri madenciliği yazılımı olan WEKA aracı [12] kullanılmıştır. Bu yüzden toplanan veriler WEKA'ya özgü dosya yapısı olan ARFF (Attribute - Relation File Format) [13] formatında düzenlenmiştir.

Sınıflandırma işleminden önce her denekten toplanan veriler üzerinde ayrı ayrı aykırı değer analizi yapılarak 110'ar veriden en aykırı olan 10'ar tanesi çıkartılmış ve her denek için 100'er girdi verisi elde kalmıştır. Aykırı değer analizi için WEKA aracında bulunan InterquartileRange yöntemi kullanılmıştır. Bu yöntem çeyrekler arası açıklığa dayanarak aykırı değerleri tespiti yapan bir filtredir [14].

3. Bulgular

Çalışmada veriler iki farklı şekilde kullanılmıştır. İlk olarak her deneğin girdiği her bir veri tekil olarak değerlendirilmiştir. İkinci olarak da bir deneğin girmiş olduğu girdi verilerinden ikisi birleştirilerek tek bir veri olarak kullanılmıştır. Belirlenen altı farklı makine öğrenmesi yöntemi ile 10 kez çapraz doğrulama yapılarak testler gerçekleştirilmiştir.

3.1. Tekil veriler ile uygulama

Tekil veriler ile gerçekleştirilen testler sonucunda Tablo 2'deki başarı oranları elde edilmiştir. En başarılı performansı gösteren Rassal Orman sınıflandırıcı ile yapılan testin sonucunda 1000 adetlik verinin 4 tanesi hatalı sınıflandırılmıştır. Bu testin sonucunda oluşan karmaşıklık matrisi Şekil 5'te sunulmuştur.

Tablo 2. Sınıflandırıcıların başarı oranları

Sınıflandırıcı	Başarı Oranları
Naive Bayes	% 94.7
Çok Katmanlı Algılayıcı	% 99.0
DVM	% 99.4
Bagging	% 97.0
Karar Ağacı	% 97.3
Rassal Orman	% 99.6
Ortalama	% 97.8

```
  a   b   c   d   e   f   g   h   i   j   <-- classified as
100   0   0   0   0   0   0   0   0   0 |   a = ibrahim
  0 100   0   0   0   0   0   0   0   0 |   b = islam
  0   0 100   0   0   0   0   0   0   0 |   c = ismail
  0   0   0 100   0   0   0   0   0   0 |   d = kemal
  0   0   0   0  98   0   0   2   0   0 |   e = kerem
  0   0   0   0   0 100   0   0   0   0 |   f = mirsat
  0   0   0   0   0   0 100   0   0   0 |   g = omer
  0   0   0   0   0   0   0 100   0   0 |   h = samet
  0   0   0   1   0   1   0   0  98   0 |   i = serdar
  0   0   0   0   0   0   0   0   0 100 |   j = yunus
```

Şekil 5. Rassal Orman testi sonucunda elde edilen karmaşıklık matrisi

Her deneğe ait 100'er adet veri ile yapılan testlerin yüksek bir başarıya sahip olduğu görülmektedir. Veri sayısındaki azalmanın başarı oranlarını nasıl etkilediğini görmek amacıyla kullanıcılardan alınan ilk 20, 40, 60 ve 80 adet veri ile testler tekrar yapılmıştır. Bu testler sonucunda elde edilen başarı oranları Tablo 3'te sunulmuştur.

Tablo 3. Farklı sayıdaki veriler ile başarı oranları

Sınıflandırıcı	Başarı Oranları (%)			
	20 adet	40 adet	60 adet	80 adet
Naive Bayes	93.5	95.75	95.167	94.750
Çok Katmanlı Algılayıcı	98.5	99.25	98.833	98.750
DVM	99.0	99.00	98.667	99.000
Bagging	95.0	96.50	97.167	97.375
Karar Ağacı	94.5	93.50	97.667	97.375
Rassal Orman	98.0	98.50	99.500	99.375
Ortalama	96.417	97.083	97.834	97.771

3.2. Birleştirilmiş veriler ile uygulama

Kullanıcıların veri girişi sırasında bazı durumlarda, derin nefes alma, etrafına göz gezdirme veya konuşma gibi çeşitli sebeplerden dolayı olağan hızlarından daha farklı davranış gösterdikleri gözlemlenmiştir. Bu yüzden, girilen bazı veriler kullanıcının ortalama değerlerinden çok farklı değerlere sahip olabilmektedir. Bu tür verilerin sınıflandırma başarısını düşürdüğü düşünülmektedir. Söz konusu durumun etkisini azaltmak ve sınıflandırma başarısını yükseltmek için veri üzerinde yeni bir önişleme daha yapılmıştır. Kullanıcının art arda yaptığı iki veri girişinin ortalaması alınarak tek bir veri haline getirilmiştir. Böylece denek başına 50'şer verilik yeni bir veri seti elde edilmiştir. Bu veri seti üzerinde yine 10 kez çapraz doğrulama yapılarak gerçekleştirilen testlerde elde edilen sonuçlar Tablo 4'te sunulmuştur.

Tablo 4. Sınıflandırıcıların başarı oranları

Sınıflandırıcı	Başarı Oranları
Naive Bayes	% 95.8
Çok Katmanlı Algılayıcı	% 99.6
DVM	% 99.8
Bagging	% 97.0
Karar Ağacı	% 97.2
Rassal Orman	% 99.6
Ortalama	% 98.2

Tablo 4'teki sonuçlara bakıldığında başarı oranları ortalamasının az da olsa artış gösterdiği görülmektedir. Bu testler sonucunda en yüksek başarıyı DVM sınıflandırıcısı göstermiştir. 500 veriden sadece 1 tanesini hatalı sınıflandıran DVM ile yapılan testin karmaşıklık matrisi Şekil 6'da sunulmaktadır.

```
  a  b  c  d  e  f  g  h  i  j   <-- classified as
 50  0  0  0  0  0  0  0  0  0 |  a = ibrahim
  0 50  0  0  0  0  0  0  0  0 |  b = islam
  0  0 50  0  0  0  0  0  0  0 |  c = ismail
  0  0  0 50  0  0  0  0  0  0 |  d = kemal
  0  0  0  0 49  0  0  1  0  0 |  e = kerem
  0  0  0  0  0 50  0  0  0  0 |  f = mirsat
  0  0  0  0  0  0 50  0  0  0 |  g = omer
  0  0  0  0  0  0  0 50  0  0 |  h = samet
  0  0  0  0  0  0  0  0 50  0 |  i = serdar
  0  0  0  0  0  0  0  0  0 50 |  j = yunus
```

Şekil 6. DVM testi sonucunda elde edilen karmaşıklık matrisi

Şekil 6'daki tek hatalı sınıflandırmanın, Şekil 5'te sunulan tekil verilerin tamamı ile ilk yapılan testteki hatalı sınıflandırmalar ile paralellik gösterdiği görülmektedir. Şekil 5'te Kerem isimli deneğe ait 2 veri Samet isimli deneğe ait olarak sınıflandırılmış, Şekil 6'da da yine Kerem isimli deneğe ait 1 veri Samet isimli deneğe ait olarak hatalı sınıflandırılmıştır. Burada, kullanıcıların ikişer verisinin birleştirilerek tek veri olarak kullanılması bu hatayı ortadan kaldırmamıştır. Ancak, Şekil 5'te Serdar isimli deneğe ait hatalı sınıflandırılan birer verinin Şekil 6'da doğru sınıflandırıldığı ve uygulanan yöntemin bu hataları ortadan kaldırdığı görülmektedir. Buradan yola çıkarak, kullanıcıların daha fazla tekil verisi birleştirilerek elde edilen yeni veri setinde daha az sınıflandırma hatası gözlemlenebileceği tahmin edilebilir.

Tablo 5. Farklı sayıdaki veriler ile başarı oranları

Sınıflandırıcı	Başarı Oranları (%)			
	10 adet	20 adet	30 adet	40 adet
Naive Bayes	95	94.5	93.333	95.50
Çok Katmanlı Algılayıcı	100	99.0	99.333	99.50
DVM	99	99.5	99.667	100
Bagging	96	95.0	97.000	95.75
Karar Ağacı	88	95.5	96.667	96.00
Rassal Orman	100	98.0	99.333	98.50
Ortalama	96.333	96.917	97.556	97.542

Kullanıcıların ikişer verisinin ortalaması ile elde edilen veriler üzerinde veri sayısının sınıflandırma başarısını nasıl etkilediğini görmek amacıyla ilk 10, 20, 30 ve 40 adet veri ile altı sınıflandırıcı tekrar test edilmiştir. Bu testlerin sonuçları Tablo 5'te sunulmuştur.

Tablo 5'de sunulan başarı oranlarına göre, test verisinin sayısındaki azalmanın DVM ve Çok Katmanlı Algılayıcı algoritmalarının başarı oranlarında genel bir artış sağladığı, diğer sınıflandırma yöntemlerinin başarı oranlarında ise bazen artış bazen düşüş gösterdiği görülmektedir.

4. Tartışma ve Sonuç

Bu çalışmada klavye kullanım stiline göre kullanıcı tanıma üzerine deneysel bir araştırma yapılmıştır. Deneylerde 10 kişilik denek grubunun 100'er adet giriş verisi kullanılmıştır. Her giriş verisi ayrı ayrı değerlendirildiğinde en başarılı sınıflandırma Rassal Orman algoritması ile %99.6 olarak elde edilirken, kullanıcıların ikişer verisi bir arada kullanıldığında en yüksek sınıflandırma performansı %99.8 olarak DVM algoritması ile elde edilmiştir. Birleştirilmiş veriler üzerinde yapılan deneylerde, sınıflandırmada kullanılan veri sayısının azaltılmasının Çok Katmanlı Algılayıcı ve Rassal Orman yöntemlerinde başarıyı düşürmediği, aksine daha da arttırdığı görülmektedir. Buradan, 10 adet kadar az sayıda birleştirilmiş veriyle bile yeni kullanıcının sisteme tanıtılabileceği ve gayet başarılı sonuçların alınabileceği gözlemlenmiştir.

Deneklerin büyük harf yazarken CAPS LOCK ve SHIFT tuşları arasında yaptıkları tercih ile rakam ve sembolleri yazarken yaptıkları tercihin kullanıcıları sınıflandırmada çok fazla önemli olduğu görülmüştür. Bu sonuç daha önceki araştırmaların sonucu ile de örtüşmektedir [15, 16]. Şekil 5'te sunulan karmaşıklık matrisinde CAPS LOCK tuşunu kullanan deneklerin SHIFT tuşu kullanan deneklerle karıştırılmadığı görülmektedir. Ayrıca, kullanıcıların her bir şifre girişini tek bir veri olarak değerlendirmek yerine, iki şifre girişini birlikte değerlendirmenin başarı oranını yükselttiği gösterilmiştir. Tablo 2 ve Tablo 4'te görüldüğü gibi, ilk yöntemde ortalama sınıflandırma başarısı %97.8 iken, birleştirilmiş verilerin kullanıldığı yöntemde ortalama sınıflandırma performansı %98.2'ye yükselmiştir. Buna göre, kritik sistemlere girişte kullanıcıdan şifresini birden fazla girmesi istenerek klavye kullanım stiline göre kullanıcı tanıma sisteminin güvenliği daha da artırılabilir.

Bu çalışmanın sonucu olarak; klavye kullanım stilinin kullanıcı tanımada ve doğrulamada güçlü bir özellik olduğu çalışmadaki yüksek başarı oranlarıyla ortaya konmuştur. Özellikle kritik öneme sahip sistemlere giriş yapılırken kullanıcıların klavye kullanım stillerinin de dikkate alınması bilgi güvenliğini artıracaktır. Gerçek kullanıcının şifresinin tahmin edilmesi, çeşitli şifre kırma programlarıyla bulunması veya başka bir şekilde yabancı bir kullanıcının eline geçmesi durumunda, eğer giriş sisteminde klavye kullanım stiline dayalı kullanıcı doğrulama yapılırsa yabancıların sisteme girişi engellenebilecektir. Yabancı kullanıcı gerçek kullanıcının şifresini bilse bile klavye kullanım stilini taklit edemeyeceğinden giriş sağlayamayacaktır. Kullanıcılara özel olarak açılan hesaplardaki verilerin gizliliği korumak ve güvenliğini arttırmak bu yolla mümkün olacaktır.

Tuş vuruş basıncını ölçebilen klavyelerin geliştirilmesiyle yeni metrikler de kullanıcı sınıflandırmada kullanılabilir. Gelecek çalışmalarda tabletler ve akıllı telefonlarda klavye kullanım stiline göre kullanıcı tanımaya yönelik araştırmaların yapılması ve sonuçların birbirleriyle kıyaslanması planlanmaktadır.

Kaynakça

[1] Titcomb, J. 2016. Do you have one of the most common passwords? They're ridiculously easy to guess. https://www.telegraph.co.uk/technology/2016/01/26/most-common-passwords-revealed---and-theyre-ridiculously-easy-to (Erişim Tarihi: 20.02.2018).

[2] Doğan, O., Aşan, H. 2016. Bireylerin Şifre Yapılarına Yönelik bir Araştırma ve Şifre Öneri Sistemi. Yönetim Bilişim Sistemleri Dergisi, 1(3), 192-201.

[3] Monrose, F., Rubin, A. D. 1999. Keystroke Dynamics as a Biometric for Authentication. Future Generation Computer Systems, 16(4), 351-359.

[4] Banerjee, S. P., Woodard, D. L. 2012. Biometric Authentication and Identification using KeystrokeDynamics: A Survey. Journal of Pattern Recognition Research, 7(2012), 116-139.

[5] Shanmugapriya, D., Padmavathi, G. 2009. A Survey of Biometric keystroke Dynamics: Approaches, Security and Challenges. International Journal of Computer Science and Information Security, 5(1), 115-119.

[6] Crawford, H. 2010. Keystroke Dynamics: Characteristics and Opportunities. Eighth Annual International Conference on Privacy, Security and Trust, 17-19 Ağustos, Ottawa, 205-212.

[7] Obaidat, M. S., Sadoun, B. 1997. Verification of Computer Users Using Keystroke Dynamics. IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, 27(2), 261-269.

[8] Robinson, J. A., Liang, V. M., Chambers, J. A. M., MacKenzie, C. L. 1998. Computer User

Verification Using Login String Keystroke Dynamics. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems ans Humans, 28(2), 236-241.

[9] Fairhurst, M., Da Costa-Abreu, M. 2011. Using keystroke dynamics for gender identification in social network environment. 4th International Conference on Imaging for Crime Detection and Prevention (ICDP), 3-4 Kasım, Londra, 1-6.

[10] Google Yardım Hesabı. 2018. Güçlü bir şifre oluşturma. https://support.google.com/accounts/answer/32040(Erişim Tarihi: 20.02.2018).

[11] Microsoft Desteği. 2018. Güçlü bir parola oluşturma. https://support.microsoft.com/tr-tr/help/4026406/microsoft-account-create-a-strong-password(Erişim Tarihi: 20.02.2018).

[12] Frank, E., Hall, M. A., Witten, I. H. 2016. Data Mining: Practical Machine Learning Tools and Techniques. Fourth Edition. Morgan Kaufmann. 654s.

[13] Anonim. 2008. Attribute-Relation File Format (ARFF). https://www.cs.waikato.ac.nz/ml/weka/arff.html(Erişim Tarihi: 20.02.2018).

[14] Purplemath. 2018. Interquartile Ranges & Outliers. http://www.purplemath.com/modules/boxwhisk3.htm (Erişim Tarihi: 20.02.2018).

[15] Lau, E., Liu, X., Xiao, C., Yu, X. 2004.Enhanced User Authentication Through Keystroke Biometrics. Teknik Rapor. Massachusetts Institute of Technology.

[16] Korkmaz, Y. 2016. Kullanıcı Giriş Sistemlerinde Yapay Sinir Ağları Kullanılarak Şifre Güvenlik Sisteminin Geliştirilmesi. Elektrik-Elektronik, Bilgisayar, Biyomedikal Mühendislikleri Bilimsel Toplantısı (EBBT), 26-27 Nisan, İstanbul, 1-4.