T.C.
AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANA BİLİM DALI

YÜKSEK LİSANS TEZİ

ÇEVRİMİÇİ SINAVLARDA TEST HİLESİNİN BELİRLENMESİNDE KULLANILAN ALTERNATİF YÖNTEMLERİN İNCELENMESİ

MUHAMMET KUMARTAŞ

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
TEZLİ YÜKSEK LİSANS PROGRAMI

Antalya, 2023

T.C.
AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANA BİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
TEZLİ YÜKSEK LİSANS PROGRAMI

ÇEVRİMİÇİ SINAVLARDA TEST HİLESİNİN BELİRLENMESİNDE KULLANILAN ALTERNATİF YÖNTEMLERİN İNCELENMESİ

YÜKSEK LİSANS TEZİ

Mhammet Kmartaş

Danışman

Doç. Dr. Bilal BarışAlkan

Antalya, 2023

DOĞRULUK BEYANI

Yüksek lisans tezi olarak sunduğum bu çalışmayı, bilimsel ahlak ve geleneklere aykırı düşecek bir yol ve yardıma başvurmaksızın yazdığımı, yararlandığım eserlerin kaynakçalardan gösterilenlerden oluştuğunu ve bu eserleri her kullanışımda alıntı yaparak yararlandığımı belirtir; bunu onurumla doğrularım. Enstitü tarafından belli bir zamana bağlı olmaksızın, tezimle ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara katlanacağımı bildiririm.

12/07/2023

Muhammet Kumartaş

AKDENİZ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Muhammet Kumartaş'ın bu çalışması 12/07/2023 tarihinde jürimiz tarafından Eğitim Bilimleri Ana Bilim Dalı Eğitimde Ölçme ve Değerlendirme Tezli Yüksek Lisans Programında Yüksek Lisans Tezi olarak oy birliği/oy çokluğu ile kabul edilmiştir.

İmza

Başkan : …

……. …….

Üye : …

……. ..……

Üye (Danışman) : ….

…… ..……

YÜKSEK LİSANS TEZİNİN ADI: ÇEVRİMİÇİ SINAVLARDA TEST HİLESİNİN BELİRLENMESİNDE KULLANILAN ALTERNATİF YÖNTEMLERİN İNCELENMESİ

ONAY: Bu tez, Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulunun …… tarihli ve …… sayılı kararıyla kabul edilmiştir.

Doç. Dr. Güçlü ŞEKERCİOĞLU
Enstitü Müdürü

TEŞEKKÜR

Test güvenliği çalışmaları ile ilgili yenilikçi çalışmalardan biri olan ve başından sonuna dek oldukça öğretici, keyifli ve verimli geçen bu tez çalışmamda; herkesten ve her şeyden evvel konu seçimimden tezin ince ayar değişikliklerine dek daim yanımda olan, ufkumu genişleten, özgür bir çalışma süreci ve ortamı sunan, hoş görülü ve sabırlı değerli danışman hocam Doç. Dr. Bilal Barış ALKAN'a, yüksek lisans eğitimim boyunca görüş ve önerileriyle bana yol gösterdiği, çalışmamın planlanıp bir ürün olarak ortaya çıkmasına kadar olan aşamayı titizlikle takip ettiği ve destek verdiği için çok teşekkür ederim.

Yüksek Lisans eğitimim boyunca bana ev sahipliği yapan güzel kardeşim Merve'ye ve çok kıymetli dostum Alper'e sonsuz teşekkür ederim.

Tüm eğitim hayatım boyunca beni koşulsuz destekleyen, her konudaki sonsuz destekleriyle bu günlere getiren, sevgi ve emeğiyle başarılarımın kaynağı olan değerli aileme çok teşekkür ederim.

Varlığı ile en büyük destekçim; her lahzada ussal ve duygusal ilhamım olan hayat arkadaşım Hilal'e çok teşekkür ederim.

ÖZET

ÇEVRİMİÇİ SINAVLARDA TEST HİLESİNİN BELİRLENMESİNDE KULLANILAN ALTERNATİF YÖNTEMLERİN İNCELENMESİ

Kumartaş, Muhammet
Yüksek Lisans, Eğitim Bilimleri Ana Bilim Dalı
Tez Danışmanı: Doç. Dr. Bilal Barış Alkan
12 Temmuz 2023, 98 sayfa

Test hilesi, çeşitli alt boyutları bulunan oldukça geniş kapsamlı bir kavramdır. Madde ön bilgisine sahip olma, kopya çekme ve gizli anlaşma ile sınav sonrası gerçekleştirilen test tahrifatı gibi alt faktörleri içeren testlerde hileli davranışlar testtin uygulanmasından önce, uygulaması sırasında veya uygulanmasından sonra haksız avantaj elde etmek için gerçekleştirilen her türlü eylem olarak tanımlanabilir.

Son yıllarda, geniş ölçekli online sınavların yaygınlaşmasıyla test hilesini belirlemeye yönelik yeni yöntemsel yaklaşımlara ihtiyaç artmaktadır. Hileli test alma davranışları özellikle psikolojik ve eğitsel değerlendirmelerde yaygın şekilde kendini gösteren bir durumdur ve test sağlayıcısının değerlendirme süreci boyunca veride saklı kalmış olabilicek bilgileri ortaya çıkarabilecek analizler yapmasını gerekli kılar. Test hilesinin belirlenmesi için literatürde önerilen pek çok yaklaşım bulunmaktadır. Fakat bu yaklaşımların tespit gücünü bir çatı altında toplayabilecek ve her bir modele dahil edilen girdi değişkenlerinin bütünsel açıdan değerlendirmeye alınmasına yönelik yapılan çalışma sayısı oldukça kısıtlıdır.

Bu çalışmada, kişi-uyum istatistikleri, kopya ve benzerlik ölçüleri, tepki süresi modelleri gibi psikometrik tabanlı yöntemlere alternatif olarak bu yöntemlerin birlikte ele alındığı yeni yaklaşımlar önerilmektedir. Bu çerçevede denetimli sınıflandırma yöntemlerinin de güçlü yönlerinin kullanılmasıyla hile davranışlarının tespit edilmesinde hatalı değerlendirme oranlarının düşürülmesinin mümkün olabileceği öngürülmektedir. Önerilen yaklaşımlar hazır bir veri seti kullanılarak analiz edilmiş ve bu yaklaşımların literatürde tanımlanan diğer klasik yöntemlere göre daha iyi hile tespit oranlarını sağladığı gösterilmiştir.

Anahtar Sözcükler: Test Hilesi, Kişi-Uyum İstatistikleri, Kopya Ve Benzerlik Analizleri, Tepki Süresi Modelleri, Denetimli Sınıflandırma Yöntemleri

ABSTRACT

EXAMINING ALTERNATIVE APPROACHES USED FOR IDENTIFYING TEST CHEATİNG IN ONLINE EXAMS

Kumartaş, Muhammet

Master of Arts, Department of Educational Sciences

Supervisor: Assoc. Prof. Dr. Bilal Barış Alkan

12 July 2023, 98 pages

Test cheating is a very comprehensive concept that includes various sub-dimensions. In tests that include sub-factors such as having prior knowledge of the item, cheating, collusion and test falsification after the test, fraudulent behavior can be defined as any action taken to gain unfair advantage before, during or after the application of the test.

In recent years, with the spread of large-scale online exams, the need for new methodological approaches to detect test cheating has increased. This study proposes new approaches as an alternative to psychometric-based methods such as person-fit statistics, copy-similarity measures, and response time models. In addition, it is thought that by using the strengths of supervised classification methods, it may be possible to reduce the rate of false evaluations in detecting cheating behaviours. It has been shown that the proposed approaches provide better cheating detection rates than other classical methods described in the literature.

In this study, new approaches are proposed as an alternative to psychometric-based methods such as person-fit statistics, duplication and similarity measures, and reaction time models, in which these methods are considered together. In this context, it is predicted that by using the strengths of the supervised classification methods, it is possible to reduce the false evaluation rates in detecting cheating behaviors. The proposed approaches have been analyzed using a ready-made data set and it has been shown that these approaches provide better fraud detection rates than other classical methods described in the literature.

Keywords: Test Cheating, Person-Fit Statistics, Copy and Similarity Analysis, Response Time Models, Supervised Classification Methods

İÇİNDEKİLER

TEŞEKKÜR i

ÖZET ii

ABSTRACT iv

İÇİNDEKİLER v

TABLOLAR LİSTESİ x

ŞEKİLLER LİSTESİ xii

KISALTMALAR LİSTESİ xiii

BÖLÜM I

GİRİŞ

1.1. Problem Durumu 1

1.2. Araştırmanın Amacı ve Problemleri 5

1.2.1. Araştırmanın Alt Problemleri 6

1.3. Araştırmanın Önemi 7

1.4. Araştırmanın Sayıltıları 9

1.5. Araştırmanın Sınırlılıkları 9

1.5. Tanımlar…….. 10

BÖLÜM II

KURAMSAL ÇERÇEVE İLGİLİ ARAŞTIRMALAR

2.1.Psikometrik Tabanlı Yaklaşımlar 12
2.1.1. Kişi-Uyum İstatistikleri 12
2.1.1.1. l_z^* Kişi-Uyum İstatistiği 15
2.1.1.2.CUSUM Kişi-Uyum İstatistiği 17
2.1.1.3. ZU3 Kişi-Uyum İstatistiği 18
2.1.2. Kopya Ve Benzerlik Ölçüleri 19
2.1.2.1. K İstatistiği 20
2.1.2.2.M4 İstatistiği 20
2.1.3. Tepki Süresi Modelleri 21
2.1.3.1. Log-Normal Tepki Süresi Modeli (LNMRT) 22
2.1.3.2. χpf İstatistiği 23
2.2. Denetimli Sınıflandırma Yöntemleri 24
2.2.1. K En Yakın Komşu Sınıflandırıcısı 25
2.2.2. Naive Bayes Sınıflandırıcısı 26
2.2.3. Rastgele Orman Sınıflandırıcısı 27

BÖLÜM III

YÖNTEM

3.1. Araştırmanın Modeli 29
3.1.1.Önerilen Yaklaşımlar 29
3.1.1.1. Yöntemlerin Birleşik Kullanımı 29
3.1.1.2. Sınav Süreci Yaklaşımı 31
3.1.1.3. Denetimli Sınıflandırma Yöntemlerinde Basit Topluluk Oyu Yaklaşımı 34
3.1.1.4. Denetimli Sınıflandırma Yöntemlerinde Birleşim Yaklaşımı 35
3.2. Araştırmanın Materyali 37
3.3. Veri Analiz Aracı 37
3.4. Veri Analiz Süreci 37
3.4.1. Kişi-Uyum İstatistiklerinin Test Hilesinin Belirlenmesinde Kullanımı 38
3.4.2. Kopya Ve Benzerlik Ölçülerinin Test Hilesinin Belirlenmesinde Kullanımı 39
3.4.3. Tepki Süresi Modellerinin Test Hilesinin Belirlenmesinde Kullanımı 39
3.4.4. Denetimli Sınıflandırma Yöntemlerinin Test Hilesinin Belirlenmesinde Kullanımı 40
3.4.4.1. Girdi Verilerinin Oluşturulması 40
3.4.4.2. Özellik Ölçekleme 41
3.4.4.3. Özellik Seçimi 42
3.4.5. Sonuç Ölçütleri 43
3.4.6. Önerilen Yöntemlerin Analizi 44
3.4.6.1. Yöntemlerin Birleşik Kullanımı İle Test Hilesinin Belirlenmesi 44

3.4.6.2. Sınav Süreci Yaklaşımı İle Test Hilesinin Belirlenmesi..... 45

3.4.6.3. Denetimli Sınıflandırma Yöntemlerinde Basit Topluluk Oyu ve Birleşim Yaklaşımı İle Test Hilesinin Belirlenmesi 46

BÖLÜM IV

BULGULAR

4.1. Kişi-Uyum İstatistikleri İle Test Hilesinin Belirlenmesine İlişkin Bulgular..... 48

4.2. Kopya Ve Benzerlik Ölçüleri İle Test Hilesinin Belirlenmesine İlişkin Bulgular 52

4.3. Tepki Süresi Modelleri İle Test Hilesinin Belirlenmesine İlişkin Bulgular 53

4.4. Denetimli Sınıflandırma Yöntemleri İle Test Hilesinin Belirlenmesine İlişkin Bulgular 54

4.5. Çalışma Kapsamında Önerilen Yaklaşımlara İlişkin Bulgular 55

4.5.1. Yöntemlerin Birleşik Kullanımı İle Test Hilesinin Belirlenmesine İlişkin Bulgular 55

4.5.2. Önerilen Sınav Süreci Modeli İle Test Hilesinin Belirlenmesine İlişkin Bulgular 57

4.5.3. Denetimli Sınıflandırma Yöntemlerinde Basit Topluluk Oyu ve Birleşim Yaklaşımı İle Test Hilesinin Belirlenmesine İlişkin Bulgular..... 58

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

5.1. Sonuç ve Tartışma..... 61

5.2. Öneriler 64

KAYNAKÇA **66**

EKLER **79**

Ek-1: Çalışmada Kullanılan R Kodları 79

ÖZGEÇMİŞ **96**

BİLDİRİM **97**

İNTİHAL RAPORU **98**

TABLOLAR LİSTESİ

Tablo 3.1. Üç Tür Test Hilesi ve Bunları Tespit Etmeye Yönelik Beş Tür İstatistiksel Yaklaşım 31

Tablo 3.2. Önerilen Sınav Süreci Yaklaşımına Göre Test Hilesi Türleri ve Bunları Tespit Etmeye Yönelik İstatistiksel Yaklaşımlar 32

Tablo 3.3. Denetimli Sınıflandırma Yöntemlerinde Test Veri Seti Çıktıları ve Etiketli Katılımcılar İçin Basit Topluluk Oyu Yöntemi (1 : Test Hilesi Yapan Bireyler ve 0: Normal Sınav Katılımcıları) 34

Tablo 3.4. Denetimli Sınıflandırma Yöntemlerinde Test Veri Seti Çıktıları ve Etiketli Katılımcılar İçin Birleşim Yaklaşımı (1 : Test Hilesi Yapan Bireyler ve 0: Normal Sınav Katılımcıları) 35

Tablo 3.5. Sonuçları Değerlendirirken Karşılaşılabilecek Durumlar 42

Tablo 4.1. Kişi-Uyum İstatistiklerinde Duyarlılık, Özgüllük ve Doğruluk 50

Tablo 4.2. Kopya ve Benzerlik Ölçülerinde Duyarlılık, Özgüllük ve Doğruluk 51

Tablo 4.3. Tepki Süresi Modellerinde Duyarlılık, Özgüllük ve Doğruluk 51

Tablo 4.4. Denetimli Sınıflandırma Yöntemleri İçin Duyarlılık, Özgüllük ve Doğruluk 52

Tablo 4.5. İki Yöntemin Birleşik Kullanımı İçin Duyarlılık, Özgüllük ve Doğruluk 53

Tablo 4.6. Üç Yöntemin Birleşik Kullanımı İçin Duyarlılık, Özgüllük ve Doğruluk 54

Tablo 4.7. Sınav Süreci Yaklaşımında Farklı Modeller İçin Duyarlılık, Özgüllük ve Doğruluk 55

Tablo 4.8. Denetimli Sınıflandırma Yöntemlerinde Basit Topluluk Oyu ve Birleşim Yaklaşımı (1 : Test Hilesi Yapan Bireyler ve 0: Normal Sınav Katılımcıları) 57

Tablo 4.9. Denetimli Sınıflandırma Yöntemlerinde Önerilen Yaklaşımlar İçin Duyarlılık, Özgüllük ve Doğruluk 57

ŞEKİLLER LİSTESİ

Şekil 3.1. Önerilen Sınav Süreci Yaklaşımı 33

Şekil 4.1. l_z^* İstatistiğinin Dağılımı, Kesme Değeri ve Hileli Davranışta Bulunan Alt Grubun Seçilmesi 47

Şekil 4.2. ZU3 İstatistiğinin Dağılımı, Kesme Değeri ve Hileli Davranışta Bulunan Alt Grubun Seçilmesi 48

Şekil 4.3. Sınav Katılımcısı 245 İçin CUSUM Süreç Çizelgesi 49

Şekil 4.4. Sınav Katılımcısı 1633 İçin CUSUM Süreç Çizelgesi 49

KISALTMALAR LİSTESİ

MTK : Madde Tepki Kuramı

KNN : K En Yakın Komşu Algoritması

NB : Naive Bayes Algoritması

RO : Rastgele Ormanlar Algoritması

LNMRT : Log-Normal Tepki Süresi Modeli

ÜS : Üst Sınır

AS : Alt Sınır

MÖ: Makine Öğrenmesi

BÖLÜM I

GİRİŞ

Bu bölümde araştırmanın problem durumu açıklanmıştır. Ayrıca araştırmanın amaç, önem, sınırlılıklar ve tanımlarına yer verilmiştir.

1.1. Problem Durumu

Eğitsel değerlendirmelerde bireylerin başarı, yetenek, performans ve tipik davranışları gibi özelliklerinin ölçülebilmesi amacıyla geliştirilen ölçme araçlarına test adı verilmektedir. Testlerden elde edilen bilgilerle doğru kararlar verilmesi ya da çeşitli çalışma ve uygulamaların etkili bir şekilde gerçekleştirilebilmesi için testlerden elde edilen ölçümlerin geçerli ve güvenilir olması gerekmektedir. Geçerlik ve güvenirlik için eğitsel ve psikolojik değerlendirme alanındaki testlerde hileli davranışlarda bulunmak birçok bağlamda giderek artan bir sorun haline gelmektedir. Testlerde hile davranışı eğitsel değerlendirme, işe alım, sertifikasyon gibi bir uygulama alanı veya eğitim kademesi, bireyin eğitim seviyesi, testin alındığı bölge gibi demografik değişkenler ile sınırlı kalmayıp her ölçütte ve durumda kendini gösterebilmektedir. Yani testlerde hileli davranışlara başvurma her yaştan öğrenci ve sınav katılımcısında görülebilmektedir. Geniş ölçekli testler, dönem sonu testleri, yerleştirme testleri, sertifika testleri gibi akademik testlerden elde edilen ölçümler kullanılarak bireyle ilgili kararlar alınmaktadır. Alınan kararlar ise bireyin hayatını etkilemektedir. Böylesi önemli durumlarda bireyin kopya çekme, madde ön bilgisinden yararlanma gibi test hilesine başvurma motivasyonu artmaktadır (Karabatsos, 2003; Meijer, 1996; Thiessen, 2008).

Test hilesi, çeşitli alt boyutları barındıran oldukça geniş kapsamlı bir kavramdır. Cizek (2012, s. 4), "bir testtin uygulanmasından önce, sırasında veya sonrasında haksız avantaj elde etmek ve yanlış sonuçlar üretmek için yapılan her türlü eylem" olarak test hilesinin tanımını yapmıştır. Test hilesi alt boyutları ile ele alınırsa üç temel bağlamda test hilesi türü vardır: madde ön bilgisine sahip olma, kopya çekme ile gizli anlaşma ve sınav sonrası gerçekleştirilen test tahrifatı (Sinharay, 2021). Madde ön bilgisine sahip olma, test maddelerinin sınav öncesi açığa çıkması ve katılımcıların maddeler hakkında değerlendirme öncesi bilgi sahibi olmasıdır.

Açığa çıkan maddeler "tehlikeye atılmış maddeler" olarak nitelendirilir. Cevapları kopyalama ve gizli anlaşmalar, test katılımcılarının sınav yönetimi esnasında yakınındaki bir başka katılımcıdan yanıtları kopyalaması ile birden fazla bireyin test maddelerini yardımlaşarak yanıtlaması olarak açıklanabilir. Test tahrifatı ise test uygulayıcıları, testi alan bireyler veya test yöneticileri tarafından değerlendirme sonrası teste verilen yanıtların değiştirilmesidir (Wollack ve diğ., 2015; Sinharay ve diğ., 2017).

Sınav öncesi, esnası ve sonrasında kontrol altında tutulamayan hile davranışları bir değerlendirme programı için olumsuz sonuçlar doğurur. Başlıca testi alanların sınav sonuçlarından elde edilen çıkarımların geçerliğini, testin ve test maddelerinin güvenirliğini olumsuz yönde etkiler. Eğitsel testlerde madde ön bilgisine sahip olma, kopya çekme, gizli anlaşmalar, şüpheli cevap değiştirme ve diğer anormal test alma davranışları testin güvenliğini, test puanlarından yapılan çıkarımların geçerliğini zayıflatmaktadır (Cizek ve Wollack, 2017; Clark ve Desharnais, 1998; Meijer, 1997; van Krimpen-Stoop ve Meijer, 2001). Bir test ile değerlendirmeye tabi olan bireylerde hileli test alma davranışlarının görülmesi yalnızca hileye başvuran test katılımcılarının değil, değerlendirmeye alınan tüm adayların sonuçlarını da etkileyeceğinden özellikle yerleştirme, setifikasyon, yetkinlik verme gibi oldukça önemli değerlendirme durumlarında ciddi endişeler ortaya çıkarmaktadır (Sinharay, 2017). Bununla birlikte hileli test alma davranışı uygulanan sınavın madde havuzunun gözden geçirilmesini, sınav prosedürlerinin yenilenmesini gerektirir. Bu işlemler test geliştiriciler için zaman ve maliyet açısından önemli bir problemdir. Ayrıca şüpheli bir test puanına dayalı olarak bir kişiye yetkinlik verilmesi, daha nitelikli bireyler yerine hileli davranış sergileyen bireylerin işe, üst öğrenim kurumları ve kurslarına alınması toplumun genel iyi oluşu başta olmak üzere daha ciddi problemlere sebep olabilir (Cizek ve Wollack, 2017). Aksi bir senaryo göz önüne alınır ise gayet normal bir sınav süreci yürütmüş ve hileye başvurmamış bir test katılımcısının hileli davrandığı yönünde bir niteleme haksız yere etiketlenen bireyde; finansal (maaş, ikramiye) kayba, fırsat ve imkanlardan (lisans eğitimi, yetkinlik kazanma, sertifika alma) yararlanamamaya, toplumsal olarak olumsuz etiketlenmeye ve hatta ruhsal sorunlara yol açabilir. Bir sınav katılımcısını hileli davrandığı yönünde etiketlemenin büyük sonuçları vardır ve adli bilişim yoluyla gerçekleştirilecek analizlerin bu hususlar göz önüne alınarak yapılmasını ve ihtiyatlı olunmasını gerekmektedir.

Test hilesinin meydana geldiğine dair bir bulgunun testi alanlar üzerindeki etkileri göz önüne alındığında, hileyi belirlemek için kullanılan bir yöntemin güçlü araştırma desteğine ve doğru tespit oranına sahip olması önemlidir. Özellikle kamu güvenliği ve etik durumların ihlal edilebileceği durumlar öncelikli olmak üzere test hilesinin belirlenmesi acil bir ihtiyaç haline gelmiştir.

Öncelikli olarak testlerde hile davranışının ortaya çıkmasını önlemek esastır ve bu sınav güvenliği çalışmalarının önemli bir kısmını kapsamaktadır. Kalan kısım değerlendirme sonrası süreci kapsar ve hileli test alma davranışları ile test hilesi yapan sınav katılımcılarını belirlemek için daha etkili yöntemler araştırılmaktadır. Bu durumu daha standart hale getirmek için Eğitsel ve Psikolojik Test Standartlarının 6.6 Standardı (Amerikan Eğitim Araştırmaları Derneği, Amerikan Psikoloji Derneği ve Eğitimde Ölçüm Ulusal Konseyi, 2014), sınavların test hilesini tespit etmek ve hile davranışının önüne geçebilmek için tanımlanmış ölçütlere sahip olması gerektiği konusunda belirli yönergeler içermektedir.

Test ortamında, test yönteminde ve nitel izleme prosedürlerinde alınacak test güvenliği önlemleri ile sınava girenlerin cevapları birbirinden kopyalama, gizli anlaşma ve somut bir materyalden doğru cevaplara erişme gibi test hilesi senaryolarının önüne geçilebilir ancak madde ön bilgisi, sınav sonrası gerçekleştirilen test tahrifatı gibi hile türleri için ayrıca nicel değerlendirmelerin yapılması gerekli görülmektedir (Meijer ve Sotaridona, 2006; Meijer, 2003). Bunun yanında yalnızca nicel yöntemler ile karar vermenin sağlıklı sonuçlar doğurmayacağı göz önüne alınarak testlerde hile tespitinde hem nitel hem de nicel yaklaşımların birlikte ele alınması daha yararlı bir yaklaşım olacaktır (Sijtsma ve Meijer, 2001; Steinkamp, 2017). Nihayetinde bir test katılımcısı için hile davranışının tespiti zor ve hassas bir değerlendirme sürecini beraberinde getirmektedir (Drasgow ve diğ., 1996; St-Onge ve diğ., 2011).

Hileli test alma davranışını belirlemenin en iyi yöntemi hakkında yaygın bir fikir birliği yoktur. Bunun yanıında hileli test alma davranışı tespitine yönelik çalışmalar ivme kazanmaktadır (Cizek ve Wollack, 2017; Drasgow ve diğ., 1996; Ferrara, 2017; Kingston ve Clark, 2014; McLeod ve diğ., 2003; Sinharay ve diğ., 2017; Sinharay ve Johnson, 2020; van der Linden ve Guo, 2008; van der Linden ve Lewis, 2015; van der Linden ve Sotaridona, 2006; van der Linden, 2009; Wang ve diğ., 2017; Wollack ve diğ., 2015; Wollack ve Fremer, 2013).

Sınavlarda farklı test hilesi türlerini belirlemek için çeşitli yöntemler önerilmektedir. Bunlar genellikle Kişi-Uyum İstatistikleri, Kopya ve Benzerlik Ölçüleri, Tepki Süresi Modelleri ve Silme Tespit İndeksleri olarak sınıflandırılmıştır (Sinharay, 2021). Her bir yöntem ile farklı düzeylerde tespit oranları ile hileli test alma davranışını belirleyebilmektedir. Fakat bu yöntem sınıflarının pratik kullanımı bir anlamda sınırlı sayılabilir, çünkü her bir yöntemin odaklandığı hile türü birbirinden farklıdır. Test hilesi çok boyutlu bir kavramdır ve klasik yöntemler ile bu kavramın yalnızca belirli bir alt kümesini oluşturan hile davranışları tespit edilebilir (Cizek ve Wollack, 2017; Kingston ve Clark, 2014; Man ve diğ., 2019; Wollack ve Fremer, 2013).

Özellikle bilgisayar tabanlı değerlendirmelerin giderek yaygın hale gelmesi pek çok sınav değişkeni ile ilgili veri elde etmeye yardımcı olmaktadır. Bu girdi bilgilerini birlikte ele almak ve aralarındaki etkileşimleri modellemek için klasik yöntemleri kullanmak zordur (Man ve diğ., 2019). Hem madde yanıtlarının hem de madde yanıt sürelerinin hileli test alma davranışlarını saptamada kullanımı giderek daha yaygın hale gelmektedir. Bunun yanında Kopya Ve Benzerlik Ölçüleri ile Kişi-Uyum İstatistiklerini hileli test alma davranışlarının belirlenmesi noktasında birlikte uygulamaya koyan çalışmalar da bulunmaktadır. Yine de bilgisayar tabanlı bir test sonrası elde edilen tüm verilerin değerlendirmeye alınarak testlerde hile davranışının belirlenmesine yönelik yapılan çalışmalar oldukça azdır (Boughton ve diğ., 2017; Fox ve Marianti, 2017; Man ve diğ., 2019; Sinharay ve Johnson, 2020; Wang ve diğ., 2017; Wollack ve Fremer, 2013).

Psikometrik tabanlı yaklaşımlar ikili madde yanıtları, kategorik madde yanıtları ve tepki süreleri gibi belirli girdilerle test güvenliği soruşturmalarını gerçekleştirmektedir. Ele alınan girdi değişkenleri de birbirinden farklıdır ve pek çok sınav değişkeni klasik yaklaşımlar ile değerlendirmeye alınmaz. Bunun yanında her yöntem birbirinden ayrı olarak kullanılır. Hileli test alma davranışını bu şekilde ele almak bu tür girdilerin tümünün hesaba katıldığı yöntemlerle ele alınmasındaki potansiyel faydayı sağlayamaz. Test hilesi yapan sınav katılımcılarının daha doğru ve etkili bir biçimde tespit edilmesinde sınav sürecinde elde edilen tüm bilgilerden nasıl yararlanacağı sorusu henüz yanıtlanmamıştır (Man ve diğ., 2019). Bunun için çeşitli model ve istatistikler önerilebilir. Bu girdi bilgilerinin birlikte kullanılması ile yöntemlerin tespit gücünün çeşitli hile davranışlarını kapsayan bir bütünlük içerisinde ele alınması faydalı bir yaklaşım olabilir. Aynı zamanda tüm tamamlayıcı bilgi parçalarını

değerlendirmeye dahil eden denetimli sınıflandırma yöntemleri bu soruya işlevsel çözümler sunabilir. Bu tür makine öğrenmesi yöntemlerinde, psikometrik tabanlı herhangi bir test hilesi belirleme yöntemi tarafından kullanılan bir değişken alt kümesinden ziyade tüm bilgilerden yararlanılarak tüm test verileri girdi değişkenleri olarak modellenir (Berkhin, 2006; Hastie ve diğ., 2009; Man ve diğ., 2019; Sinharay, 2020; van der Linden ve Guo, 2008). Bu bağlamda yöntemlerin tespit gücünü çeşitli hile davranışlarını kapsayan bir bütünlük içerisinde ele almak hiç şüphesiz faydalı bir yaklaşım olacaktır (van der Linden ve Guo, 2008).

Bu çalışmada, testlerde hileli davranışların belirlenmesi için literatürde yer alan yöntemlerin etkinliğini araştırmak, birden fazla yöntemin bir test senaryosunda hileli davranış tespiti için kullanımının kullanışlılığını denetlemek, literatürde kullanılan yöntemlerden bazılarının bir hile türü ile sınırlandırıldığı varsayımıyla bütünsel bir süreç modeli içerisinde değerlendirerek sınav güvenliğine yapabileceği katkıyı araştırmak, çeşitli denetimli makine öğrenmesi yöntemlerinin ve bu yöntemlerin birlikte ele alınmasının test hilesini belirlemedeki etkililiğini test yönetimi boyunca elde edilen pek çok değişkeni sisteme dahil ederek en iyi şekilde belirlemek amaçlanmıştır. Ayrıca çalışmada, hileli test alma davranışları ile ilgili iki tamamlayıcı bakış açısı sunmak hedeflenmektedir. İlki, şaibeli sınavlarda gözlemsel ve nitel yöntemler ile incelenecek optimum alt grubun belirlenmesi; ikincisi ise nitel yöntemler ile hakkında hileli davranış göstergesi tespit edilmiş sınav katılımcılarının hileli davrandığına yönelik istatistiksel kanıtlar sunmak için çeşitli yaklaşımları ele almaktır. Açımlayıcı ve doğrulayıcı olarak tanımlanabilecek iki ayrı bakış açısı yoluyla, ilgili araştırmalarda sıklıkla kullanılan gerçek bir veri seti üzerinden çok boyutlu bir bakış açısı sunulmaya çalışılmıştır.

1.2. Araştırmanın Amacı ve Problemleri

Bu araştırmanın genel amacı, psikometrik tabanlı çeşitli göstergelerin, birden fazla göstergenin birlikte kullanımının, önerilen sınav süreci modelinin, denetimli sınıflandırma yöntemlerinin ve denetimli sınıflandırma yöntemlerinden elde edilen çıktılarının bütünsel kulllanımının hileli test alma davranışını belirleme üzerindeki etkililiğini araştırmaktır. Böylece giderek daha yaygın bir biçimde alanda araştırılan testlerde hileli davranışların belirlenmesi için çok yönlü ve tespit doğruluğu yüksek olan modellerin ortaya koyulması amaçlanmıştır.

1.2.1. Araştırmanın Alt Problemleri

Araştırmanın amacı doğrultusunda yanıt aranan araştırma problemleri;

1) Çalışmada kullanılan veri seti için hileli test alma davranışı sergileyen sınav katılımcılarını doğru tespit etme oranları (duyarlılık, özgüllük, doğruluk);

   1.1) Kişi-Uyum İstatistikleri olan l_z^*, CUSUM ve ZU3 istatistikleri için nasıldır?

   1.2) Kopya ve Benzerlik Ölçüleri olan K ve M4 istatistikleri için nasıldır?

   1.3) Tepki Süresi Modelleri olan LNMRT ve χpf istatistikleri için nasıldır?

2) Birden fazla Kişi-Uyum İstatistiği, Kopya ve Benzerlik Ölçüsü ve Tepki Süresi Modelinden kanıt toplamanın hileli test alma davranışı sergileyen sınav katılımcılarını doğru tespit etme oranları (duyarlılık, özgüllük, doğruluk);

   2.1) Kişi-Uyum İstatistikleri ile Kopya ve Benzerlik Ölçüsü analizlerinin birleşik kullanımı durumunda nasıldır?

   2.2) Kişi-Uyum İstatistikleri ile Tepki Süresi Modellerinin birleşik kullanımı durumunda nasıldır?

   2.3) Tepki Süresi Modelleri ile Kopya ve Benzerlik Ölçüsü analizlerini birleşik kullanımı durumunda nasıldır?

   2.4) Kopya ve Benzerlik Ölçüleri, Kişi-Uyum İstatistikleri ve Tepki Süresi Modellerinin analizlerinin birleşik kullanımı durumunda nasıldır?

3) Önerilen Sınav Süreci Yaklaşımı ile kullanılan veri seti için hileli test alma davranışı sergileyen sınav katılımcılarının doğru tespit edilme oranları (duyarlılık, özgüllük, doğruluk) çeşitli model kombinasyonları altında nasıldır?

   3.1) Önerilen Sınav Süreci Yaklaşımı için hileli test alma davranışı sergileyen sınav katılımcılarının tespit edilmesinin yanında test hilesinin gerçekleştiği test merkezlerinin doğru tespit edilme oranı nasıldır?

4) Denetimli Sınıflandırma Yöntemlerinin hileli test alma davranışı sergileyen sınav

katılımcılarını doğru tespit etme oranları (duyarlılık, özgüllük, doğruluk);

4.1) K-En Yakın Komşu Sınıflandırma Algoritması için nasıldır?

4.2) Naive Bayes Sınıflandırma Algoritması için nasıldır?

4.3) Rastgele Ormanlar Sınıflandırma Algoritması için nasıldır?

5) Denetimli sınıflandırma yöntemlerinin çıktılarından yola çıkarak önerilen yaklaşımların hileli test alma davranışı sergileyen sınav katılımcılarının doğru tespit edilme oranları (duyarlılık, özgüllük, doğruluk);

5.1) Önerilen Basit Topluluk Oyu Yaklaşımı için nasıldır?

5.2) Önerilen Birleşim Yaklaşımı için nasıldır?

1.3. Araştırmanın Önemi

Problem durumu bölümünde de kısmen bahsedildiği üzere hileli test alma davranışlarının belirlenebilmesi amacıyla birbirinden farklı pek çok yöntem ve istatistik geliştirilmiştir. Test güvenliği araştırmalarında genellikle tek bir yöntem sınıfına, girdi olarak yalnızca tek bir değişken türüne veya hile davranışının tek bir boyutuna odaklanıldığı görülmektedir (Anderson, 1984; Armstrong ve Shi, 2009; Bolsinova ve Tijmstra, 2018; Boughton ve diğ., 2017; Bradlow ve diğ., 1998; Cizek ve Wollack, 2017; Drasgow ve diğ., 1985; Drasgow ve diğ., 1991; Emons ve diğ., 2002; Fox ve Marianti, 2017; Glass ve van der Linden, 2010; Gutmann, 1944; Hanson ve diğ., 1987; Harnisch, 1983; Hauser ve diğ., 2011; Holland, 1996; Karabatsos, 2003; Marianti ve diğ., 2014; Maynes, 2014; Maynes, 2017; Meijer ve Sijtsma, 2001; Meijer ve Tendeiro, 2012; Meijer, 2003; Molenaar ve Hoijtink, 1990; Molenaar, 1990; Montgomery, 2012; Nering, 1995; Qian ve diğ., 2016; Reise, 1990; Reise; 1995; Schnikpe ve diğ., 1999; Schnipke ve Scrams, 1999; Sinharay, 2017; Sinharay, 2018; Sinharay, 2020; Snijders, 2001; Sotaridona ve Meijer, 2002; Sotaridona ve Meijer, 2003; Tatsuoka ve Tatsuoka, 1983; Tenderio ve diğ., 2016; van der Flier, 1977; van der Flier, 1980; van der Flier, 1982; van der Linden ve Guo; 2008; van der Linden ve Sotaridona, 2006; van der Linden, 2007; van der Linden, 2009; van Krimpen-Stoop ve Meijer, 1999; van Krimpen-Stoop ve Meijer, 2001; Wang ve Xu, 2015; Wise ve diğ., 2010; Wollack ve Cohen, 1998; Wollack ve

diğ., 2015; Wollack, 1996; Wollack, 2003; Wollack, 2006; Zopluoglu, 2019).

Test hilesinin belirlenmesinde özellikle kritik ve yüksek risk düzeyine sahip koşullar göz önüne alındığında optimum tespit gücünün nasıl sağlanacağı sorusu henüz etkili bir şekilde aydınlatılmamıştır. Halihazırda hileli test alma davranışlarının tespitinde ilgili çalışmaların salt çoğunluğunun istatistiklerin türetilmesi ve geliştirilmesi ile anormal yanıt kalıplarını belirleme oranına odaklandığını belirtilmektedir (Karabatsos, 2003; Meijer ve Sijtsma, 2001; Meijer, 2003). Buna alternatif olarak yöntemlerin gücünü ve sınav öncesinde, esnasında ve sonrasında elde edilen tüm bilgileri analizlere dahil ederek test verilerinden en verimli biçimde yararlanmaya odaklanmak, tespit doğruluğu yüksek ve çok yönlü çıkarımlar sağlamaya yardımcı olabilir. Gerçekleştirilen araştırmada bu nokta göz önünde bulundurularak hileli test alma davranışlarının tespitinde daha etkili ve daha yüksek tespit gücüne sahip olacak bir takım bütünsel ve birleşik yöntemler gerçek bir veri seti ile irdelenmiştir. Bu bağlamda mevcut çalışma birbirinden farklı istatistiksel yöntem sınıflarında bulunan ve farklı girdiler ile test hilesini tespit etme yaklaşımları sunması açısından kullanışlı bir bakış açısı sağlayabilir ve daha yüksek tespit oranları sunma noktasında faydalı olabilir.

Sınav yöneticileri ve sınav sonrası hile tespiti ile ilgili soruşturma gerçekleştirecek araştırmacılar için mevcut çalışmanın odağında olan çift yönlü bakış açısı daha operasyonel ve pratik bir uygulama mantığı ortaya koyabilir. Bu bir test senaryosunda, hileye ilişkin bir kaygı olması halinde çok yüksek sayılarda sınav katılımcısının gözden geçirilmesi ve soruşturmaya dahil edilmesinin mümkün olmadığı veya maliyetinin oldukça yüksek olduğu durumlarda şaibeli sınav için çalışmada önerilen yöntemlerin bir kısmı ile incelenecek sınav alt grubunun belirlenmesi; diğer taraftan halihazırda hakkında nitel ve gözlemsel yöntemler ile hile tespiti gerçekleştirilen bireyler hakkında etiketlemeyi sağlamlaştırmak ve gerekli önlemleri almak için nicel kanıt elde etme yönleri ile uygulayıcılar için daha işlevsel bir yönlendirme sağlayabilir.

Çalışmada kullanılan yaklaşımlar testlerde hileli davranışları birey düzeyinde ele almaktadır. Ancak önerilen yaklaşımlardan biri olan sınav süreci modelinin bir getirisi de özellikle hileli davranışlarda bulunulmuş olabilecek sınav merkezlerinin belirlenmesi açısından grup veya merkezler düzeyinde de hile tespiti için kullanılan yöntemlere alternatif sağlayabileceği görüşüdür (Wollack ve Eckerly, 2017; Maynes, 2014; Sinharay, 2018). Bu açıdan hesaplama maliyetinden tasarruf ederek uygulayıcılara ve araştırmacılara pratik bir görüş sağlayabilir.

Eğitsel ölçümlerde ve özellikle çalışmanın konu alanı olan testlerde hile davranışlarının incelenmesinde makine öğrenmesi yöntemlerinin kullanımı az sayıda araştırma ile sınırlıdır (Man ve diğ., 2019; Wang ve diğ., 2019; Zopluoglu, 2019). Bunun yanında eğitimde makine öğrenmesi algoritmalarının kullanımı giderek daha yaygın hale gelmektedir. Çalışmanın odak noktası olan değerlendirme sonrası adli bilişim yoluyla test hilesinin tespiti noktasında makine öğrenmesi yöntemlerinin kullanılmasının çeşitlendirilmesi ve farklı ölçülerle türetilmiş değişkenlerin de sürece dahil edilmesi açısından bu araştırmanın alandaki çalışmalara çeşitlilik kazandıracağı da düşünülmektedir.

Yine benzer şekilde test hilesi araştırmalarında ve daha genel bir çerçevede eğitim bilimlerini ve sosyal bilimleri çatısı altına alan disiplin alanlarında makine öğrenmesi yöntemleri yer yer kullanılsa da makine öğrenmesi algoritmalarının birleşik ve bütünsel gücünden yararlanma üzerine henüz gerçekleştirilmiş bir çalışma bulunmamaktadır. Bu yöntemler temel olarak; algoritma artırma (boosting algorithms), torbalama algoritmaları (bagging algorithms) ve istifleme algoritmalarıdır (stacking algorithms). Çalışma kapsmında önerilen yaklaşımlardan basit topluluk oyu ve birleşik kullanım modelleri oldukça sade bir bakış açısı kullanarak hile belirlemede tespit gücünün artırılması konusunda ümit vaat edici olabilir.

1.4. Araştırmanın Sayıltıları

Araştırmada kullanılan veri setinin veri sağlayıcı tarafından şeffaf bir biçimde paylaşıldığı, söz konusu Yeterlik Belgesi veri setinin bilgisayar tabanlı testler ile elde edilebilecek tüm girdi değişkenlerini kapsadığı ve çalışma materyalinin ilgili araştırmalarda da sıklıkla kullanılan ve başvurulan kaynaklardan biri olması sebebiyle gerçekleştirilecek çalışma için uygun olduğu varsayılmıştır. Analizler gerçekleştirilirken seçilen parametrelerin ve ölçütlerin (kesme değerleri, anlamlılık düzeyi, algoritma durdurma kuralları, özellik ölçekleme ve özellik seçimi yöntemi gibi) literatür ve çalışma materyalinin etkili şekilde değerlendirilmesi göz önünde tutularak optimum tercihler olduğu varsayılmıştır.

1.5. Araştırmanın Sınırlılıkları

Araştırma Kişi-Uyum İstatistiklerinde l_z^*, ZU3 ve CUSUM İstatistikleri, Kopya Ve

Benzerlik Ölçülerinde K ve M4 İndeksleri, Tepki Süresi Modellerinde LNMRT ve χpf İstatisitikleri ile Denetimli Sınıflandırma Yöntemlerinde KNN Sınıflandırıcısı, Naive Bayes Sınıflandırıcısı, Rastgele Ormanlar Sınıflandırıcısı algoritmaları ve kullanılan yöntemler içinde tercih edilen ölçüler ve algoritma durdurma kuralları ile sınırlandırılmıştır.

1.6. Tanımlar

- **Test Hilesi:** Testin uygulaması öncesi, esnası ve sonrasında yapılan, kişilere haksız avantaj sağlayan eylem ve durumlardır. Bir bireyin gerçek bilgi ve becerisini temsil etmeyen bir puan için kasıtlı olarak başvurduğu test alma davranışlarıdır.

- **Kişi-Uyum İstatistikleri:** Testi alan bireylerden elde edilen yanıt örüntülerinin istatistiksel bir ilgi modeli altında kalan gruptan gelen yanıt kalıplarından anlamlı düzeyde farklılaşıp farklılaşmadığını araştıran istatistiklerdir. Anormal tepki kalıpları olan bireyleri belirlemek için geliştirilmiştir.

- **Kopya ve Benzerlik Ölçüleri:** Test katılımcıları arasındaki eşleşen madde örüntülerini ele alarak farklı yaklaşımlar ile doğru-yanlış eşleşmeleri niceliksel olarak tespit edip anlamlı ölçüde benzerlik gösteren katılımcıları belirlemeyi amaçlayan istatistiklerdir.

- **Tepki Süresi:** Bir sınav katılımcısının testteki bir maddeye yanıt verirken kullandığı süredir (Kingston ve Clarke, 2014).

- **Tepki Süresi Modelleri:** Test katılımcılarının maddelere verdikleri tepki sürelerine ilişkin verilere dayanan gizil değişkenli yaklaşımlardır. Bu modellerin kullanımıyla test hilesinin belirlenmesinde pek çok yaklaşım önerilmiştir.

- **Makine Öğrenmesi:** Hesaplama performansı artırmak veya doğru tahminlere sahip olmak için geçmiş bilgileri kullanan hesaplama yöntemleridir. Sınıflandırma, kümeleme, regresyon ve boyut azaltma gibi çeşitli alanlarda uygulanır. Makine öğrenmesi algoritmaları denetimli ve denetimsiz olarak kategorize edilebilir.

- **Denetimli Sınıflandırma Yöntemleri:** Algoritmanın girdi değişken değerleri ile çıktı değişkenlerini sınıflandırmada kullanılabilecek bir tahmin fonksiyonu üretmek için girdi ve çıktı değişkenlerini içeren bir eğitim veri setinden öğrenmesini gerektirir. Yeni birimler

daha sonra tahmin fonksiyonuna göre sınıflandırılır.

- **Ölçekleme (Normalizasyon):** Denetimli makine öğrenmesi yöntemlerinde verileri sınıflandırmaya hazır hale getirmek için girdi değişkenlerini ortak bir ölçeğe aktarma işlemidir.

- **Özellik Seçimi:** Bir denetimli sınıflandırma yönteminde modelden çıkarımlar yapmak için girdi değişkenleri arasından daha az sayıda olan bağımsız değişkenlerin seçilmesi gerekmektedir. Özellik seçimi, model yapımında kullanılacak en tutarlı, yedeksiz ve ilgili özellikleri ayırma işlemidir. Veri kümelerinin boyutu ve çeşitliliği artmaya devam ettiğinden, veri kümelerinin boyutunu düzenli olarak azaltmak önemlidir. Özellik seçiminin temel amacı, tahmine dayalı bir modelin performansını iyileştirmek ve modellemenin hesaplama maliyetini azaltmaktır

- **R Programı:** İstatistiksel analiz sürecinde sıklıkla kullanılan açık kaynak kodlu bir programdır.

BÖLÜM II

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde test hilesini belirleme yöntemlerine kısaca değinilmiştir. İlk bölümlerde ele alınan Psikometrik Tabanlı Yöntemler içinde çalışma kapsamına alınan Kişi-Uyum İstatistikleri ve bunlardan l_z^*, CUSUM ve ZU3 Kişi-Uyum İstatistiklerine; Kopya Ve Benzerlik Ölçüleri ile bu ölçülerden K ve M4 İndekslerine; Tepki Süresi Modelleri ve çalışmada kullanılan Tepki Süresi Modelleri olan LNMRT ve χpf İstatistiklerine değinilmiştir. Bunun yanında ikinci kısım olan Denetimli Sınıflandırma Yöntemleri için araştırmada kullanılan K-En Yakın Komşu Sınıflandırma Algoritması, Naive Bayes Sınıflandırma Algoritması ve Rastgele Orman Sınıflandırma Algoritmasından bahsedilmiştir.

2.1. Psikometrik Tabanlı Yaklaşımlar

Psikometrik tabanlı yaklaşımlar; Kişi-Uyum İstatistiklerini, Kopya Ve Benzerlik Ölçülerini, Tepki Süresi Modellerini, Puan Farklarının Değerlendirilmesini, Silmelerin Ve Cevap Değişikliklerinin Analizini kapsamaktadır (Sinharay, 2020). Bu yaklaşımlar hileli test alma davranışlarını tespit etmek için madde yanıtlarını, madde yanıt sürelerini veya her ikisini birden kullanır. Araştırma kapsamına alanyazında oldukça popüler olan ilk üç yöntem sınıfı alınmıştır ve bu yöntem sınıflarında tercih edilen göstergeler farklı girdi değişkenleri ile istatistiksel temelleri açısından birbirinden ayrı bakış açıları sağlayacak şekilde seçilmiştir.

2.1.1. Kişi-Uyum İstatistikleri

Kişi-Uyum İstatistikleri eğitsel psikometri başta olmak üzere çeşitli disiplinlerde giderek daha fazla ilgi görmektedir. Karabatsos (2003) ile Meijer ve Sijtsma (2001) çok sayıda Kişi-Uyum İstatistiğinin kapsamlı incelemesini sağlamışlardır. Meijer (2003) kişi-uyum istatistikleri ile ilgili çalışmaların salt çoğunluğunun istatistiklerin türetilmesi ve geliştirilmesi ile anormal yanıt kalıplarını belirleme oranına odaklandığını belirtmiştir.

Kişi-Uyum İstatistikleri, test katılımcılarının maddelere verdikleri cevap örüntülerinin bir istatistiksel model altında kendilerinden beklenen yanıt kalıpları ile ne düzeyde uyuştuğunu

inceler, herhangi bir farklılaşma veya sapma olup olmadığını araştırır. Madde Tepki Kuramı tabanlı istatistiklerde hile davranışının göstergesi kabul edilen cevap örüntüleri, herhangi bir Madde Tepki Kuramınca öngörülen yanıt modelinden anlamlı düzeyde farklılaşan yanıt kalıplarıdır. Bir bireyin kendisinden beklenen yetenek düzeyinden daha yüksek zorlukta soruları doğru yanıtlaması veya yetenek düzeyi göz önüne alındığında kendi normunda kolay sayılabilecek sorulara yanlış yanıtlar vermesi anormal tepkiler olarak düşünülebilir (Levine ve Drasgow, 1988; Meijer ve Sijtsma, 2001; Reise, 1990; Reise, 1995; Reise ve Widaman, 1999; van Krimpen-Stoop ve Meijer, 1999). Anormal yanıt örüntülerinin tespiti, bir sınav katılımcısının tepkilerinin bir MTK modeli ile ne düzeyde uyduğunu değerlendirerek çeşitli istatistiksel yöntemler (Kişi-Uyum İstatistikleri) ile gerçekleştirilir. Kişi-Uyum İstatistikleri çoğunlukla MTK modellerinin varsayımları, ölçülen özelliğin tek boyutluluğu, birimlerin yerel bağımsızlığı gibi MTK ilkeleri kullanılarak türetilmiştir. Kişi-Uyum İstatistikleri katılımcıların yanıt kalıplarının bir MTK modeli ile ne kadar tutarlı olduğunun ölçüsüdür. Bu istatistikler bireyin test alma davranışlarında yetenek kestirici davranışlar harici etmenlerin etkili olup olmadığını belirlemede kullanılır ve yine bireylerin yetenek düzeyleri göz önüne alınarak benzer yetenek düzeyinde olan bireylerin yanıt örüntülerinin farklılaşmasına odaklanır (Harnisch, 1983; Meijer, 2003; Reise, 1990).

Kısaca Kişi-Uyum İstatistikleri;

-Sınav katılımcısının yanıt kalıplarının bir MTK modeli ile uyumluluğunu değerlendirme,

-Testi alanın yetenek tahmininin gerçek yetenek düzeyini yansıtıp yansıtmadığını araştırma,

-Anormal yanıt kalıplarını belirleme ve beklenmeyen yanıtların nedenleri için gerekçe sağlama,

-Değerlendirme noktasında yetersiz olan testleri ve test içeriğinde tehlikeye atılmış maddeleri belirleme,

-Değerlendirmenin yapıldığı yetenek ve yapısal durumların uygunluğunu inceleme,

amaçları etrafında başlıca modele uymayan yanıt kalıplarının belirlenmesi olmakla birlikte, birey uyumunu değerlendirmektedir (Ayala, 2009; Karabatsos, 2003).

Bununla birlikte, modele uymayan yanıt örüntüleri pek çok nedenle bağdaştırılabilir ve bunların tamamı hileli test alma davranışının göstergesi olarak kabul edilemez. Cevapları kopyalama, gizli anlaşma, madde ön bilgisi gibi test hilesi türleri başta olmak üzere anormal yanıt kalıplarına düşük motivasyon, dikkatsizlik, rastgele tahminlerde bulunma gibi çeşitli durumlar neden olabilir (Belov, 2013; Tatsuka, 1984). Bu tür durumlar da yetenek tahminlerini etkileyebilir ve gerçek beceri düzeyi ile ilgili ciddi sapmalara neden olabilir. Bireylerin madde yanıt kalıpları böylesi durumlarda potansiyel yetenek düzeylerini yansıtmaz. Anormal yanıt kalıplarına sahip olan test katılımcıları kendileri açısından riskli sonuçlar doğurabilecek puanlar ve değerlendirme sonuçları ile karşı karşıya kalabilirler (De La Torre ve Deng, 2008). Belov'a (2013) göre test puanlarının geçerliğinin ve güvenirliğinin sağlanması ve maddelerin operasyonel hale getirilmesi için modele uymayan yanıt örüntülerinin testten çıkarılması gerekmektedir.

Test hilesi konusunda değerlendirme yapan araştırmacılar ve test yöneticileri, pek çok Kişi-Uyum İstatistiğine erişim sağlayabilir ve ilgili test senaryosu için en uygun olanı seçmekle yükümlüdürler. Geliştirilen ve iyi araştırılmış çok sayıda Kişi-Uyum İstatistiği mevcuttur. Hangi Kişi-Uyum indeksinin uygun olduğu ve daha iyi tespit oranları sağlayacağı gerçekleştirilen sınavın yapısına ve uygulama sürecine bağlı olarak değerlendirilir. Soru tipi, testin uzunluğu, kullanılan MTK modeli, madde ve test istatistikleri gibi faktörler kullanılacak Kişi-Uyum İstatistiğinin seçilmesinde belirleyicidir. Örneğin, test uzunluğu ve madde güçlükleri heterojen olan sınavlarda Kişi-Uyum İstatistiklerinin daha doğru tespit oranları ortaya koyduğu düşünülmektedir, soru sayısının artması ve güçlük düzeyi daha yüksek soruların ağırlıkta olması uyumsuz yanıt modellerinin tespitini kolaylaştırır (St-Onge ve diğ., 2011; Melenaar ve Hoijtink, 1990, Molenaar ve Hoijtink, 1996). Bu nedenle test uygulayıcıları ve sınav sonrası adli bilişim yoluyla değerlendirme yapacak araştırmacılar bu prosedürlerin en iyi şekilde nasıl uygulanacağı konusunda bilgili olmalı, önceki uygulama ve araştırmalardan yararlanmalıdır.

Kişi-Uyum İstatistiklerinin başlıca sınırlı olduğu durum, yanıt kalıplarının beklenen modelden farklılaştığını belirlemesin yanında söz konusu sapmanın nedeni hakkında bir açıklama getirememesidir. Test hilesi bağlamında hileli davranış türlerinin hangi Kişi-Uyum İstatistikleri ile daha doğru tespit oranları verdiği incelenmiştir. Hileli test alma davranışının saptanması, rastgele yanıt verme ve düşük motivasyon sebebiyle uyumsuz yanıt modelinin

ortaya çıkmasını tespit etmekten daha güçtür ve alınan kararın önemi daha ciddi sonuçlar doğurmaktadır (Karabatsos, 2003; Meijer ve Sijtsma, 2001). Bir diğer husus, bir yanıt kalıbındaki sapma miktarı ile hileli davranış göstergesi olması arasında doğrusal bir ilişki olmayabileceğidir. Sapma düzeyinin belirli bir eşiği aşması doğru tespit oranında bir azalmaya neden olabilir, bunun nedeni bir yanıt deseni yüksek düzeyde sapmaya sahip olduğunda yetenek kestiriminin de bu durumdan etkilenmesi ve kişi-uyum istatistiğinin etkinliğinin düşmesidir (St-Onge ve diğ., 2011). Ayrıca kullanılan istatistiğin ampirik dağılımı hileli davranışı tespit noktasında belirleyicidir ve dağılımı belirlemek için ileri düzeyde uzmanlık bilgisi gerektirir. Söz konusu ampirik dağılım teorik dağılım ile örtüşmez ise istatistiğin hileli davranmış olabilecek bireyler ile normal test katılımcılarını ayırt etme yeteneği azalacaktır.

Model uyumuna dayanan Kişi-Uyum analizlerinin sonuçları hem katılımcı hem de sınav yöneticileri için oldukça ciddi sonuçlar doğurabileceği ve maliyetli olacağı için alınan kararları destekleyecek çıktıların artırılması ve daha fazla kanıtın nitel ve gözlemsel önlemler ile birlikte kullanılması yararlı olacaktır.

Kişi-uyum istatistiklerinden l_z^*, CUSUM ve ZU3 çalışmada tercih edilen göstergeler olmuştur. Bu istatistikler yüksek tespit oranları gösterdikleri, istatistiksel temelleri açısından farklı bakış açıları ve tespit çeşitliliği sağlayacakları düşünülerek araştırma bağlamında kullanılmıştır (Armstrong ve Shi, 2009; Bradlow ve diğ., 1998; Emons ve diğ, 2002; Montgomery, 2012; Reise, 1995; Snijders, 2001; van der Flier, 1982).

2.1.1.1. l_z^* Kişi-Uyum İstatistiği

l_0 olasılık fonksiyonunda, testi alan bireyin ikili yanıtının olasılık dağılımı bir Bernoulli dağılımını takip etmektedir ve rassal değişken iki değer almaktadır. Bu Bernoulli rassal değişkenin logaritması alınarak herhangi bir Madde Tepki Kuramı modelince bireyden beklenen yanıtları gözlemlemenin ne kadar olası olduğunun bir göstergesi olduğu kabul edilir. Bu göstergenin değerinin büyük negatif olduğu (mutlak değerce arttığı) durumlar katılımcının cevap örüntülerinin MTK modeli altında tahmin edilenden farklılaştığını göstermektedir. l_0 olasılık fonksiyonu;

$$l_0 = \sum_{n=1}^{N} (x_i ln P_i(\theta) + (1 - x_i) \ln[1 - P_i(\theta)]) \quad (1)$$

şeklinde ifade edilir (Levine ve Rubin, 1979). Burada N, test katılımcılarını; x_i, incelenen test katılımcısının i maddesinde puanlanmış yanıtını (doğru ise 1, yanlış ise 0); $P_i(\theta)$, test katılımcısının yetenek düzeyini temsil etmektedir. Test hilesinin tespiti ve ilgili diğer araştırmalar için l_0 olasılık fonksiyonunun kullanımını sınırlayan noktalar; istatistiğin standardize olmaması ve bir yanıt kalıbını anormal olarak etiketleyebilmek için yokluk hipotezine ihtiyaç duymasıdır. l_z İstatistiği l_0 olasılık fonksiyonunun sınırlamalarını ortadan kaldırmak için standartlaştırılmış halidir (Cizek ve Wollack, 2017; Drasgow ve diğ., 1985; Levine ve Rubin, 1979; Meijer ve Sijtsma, 2001). l_z İstatistiğinin eşitliği şu şekilde ifade edilebilir (Drasgow ve diğ., 1985):

$$l_z = \frac{l_0 - \mathrm{E}(l_0)}{[Var(l_0)]^{1/2}} \tag{2}$$

Olasılık temelli parametrik l_z İstatistiği ve türevleri alanda en çok kullanılan kişi-uyum analizlerinin başında gelmektedir. Bu indeksin popülerliği başlıca iki nedenden kaynaklanmakatadır; (1) hesaplamasının kolay olması, (2) asimptomatik dağılımının standart bir normal dağılımı takip etmesinden dolayı dağılımın kolay yorumlanabilirliği. Bu nedenlerden dolayı değerlendirmeye alınan bireylerin l_z puanları normal dağılım eğrisindeki z puanları gibi düşünülerek belirli bir standart puanın üzerindeki l_z değerleri uç değerler olarak kabul edilip hileli test alma davranışının göstergesi olarak düşünülebilir. Bu istatistik yapılan bir çok araştırmada iyi tespit oranları sağlamış, anormal yanıt örüntülerini belirleme noktasında oldukça iyi çıktılar veren bir indeks olduğu konusunda yaygın bir görüş benimsenmiştir (Armstrong ve diğ., 2007; Cizek ve Wollack, 2017; Drasgow ve diğ., 1991; Li ve Ojenik, 1997, Rupp, 2013).

l_z İstatistiğinin anormal tepki modellerini belirlemede başarılı olması için l_z puanarının normal olarak dağılması gerekir. Fakat l_z'nin teorik dağılımını gerçek veriler kullanarak elde etmek çok rastlanır bir durum değildir. Testin uzunluğu, madde istatistiklerinin dağılımı, yetenek parametrelerinin tahmini veya gerçek değerinin hesaplamaya dahil edilmesi gibi faktörler istatistiğin dağılımı üzerinde belirleyici olmaktadır (Cizek ve Wollack, 2017; Drasgow ve diğ., 1985; Molenaar ve Hoijtink, 1990; Seo ve Weiss, 2013; van Krimpen Stoop ve Meijer, 1999). Ayrıca l_z'nin standart bir normal dağılımı izlediği varsayılsa da Nering (1995), Molenaar ve Hoijtink (1990) ve Reise (1995) gerçekleştirdikleri çalışmalarda tahmini yetenek

değerlerinin kullanılması halinde bu varsayımın geçerli olamayabileceğini açıklamışlardır. Bu probleme çözüm olarak yetenek kestirimleri ile asimptomatik standart normal dağılımı elde etmek için l_z İstatistiğinin bir varyantı önerilmiştir (Snijders, 2001). Bu istatistik, eşitlik (3) ile verilmiştir (Snijders, 2001).

$$l_z^* = \frac{l_0(\theta) - \mathrm{E}\big(l_0(\theta)\big) + c_j(\theta) r_0(\theta)}{\sqrt{V}(l_0(\theta))} \tag{3}$$

Denklemde ifade edilen $l_0(\theta)$, olasılık yoğunluk fonksiyonundan elde edilen parametreyi; c_j, rastgele tahminde bulunma ile ilgili olan şans parametresini; r_0 ise beklenen standart kalıntıyı ifade etmektedir. Bu istatistiğin büyük negatif değerleri uyumsuzluğu, büyük pozitif değerleri aşırı uyumu göstermektedir.

2.1.1.2. CUSUM Kişi-Uyum İstatistiği

CUSUM yaygın olarak kullanılan kişi-uyum istatistiklerinden farklıdır. Halihazırda kullanılan kişi-uyum istatistiklerinin sınırlı olduğu nokta analizlerde madde sıralama bilgisini kullanmamalarıdır. Cevaplara yalnızca toplu olarak bakıldığında, testin farklı bölümlerinde birbirinin aksi yönde olan sapmalar anormal yanıt kalıplarının tespit edilememesine yol açabilir. Bu durumda peş peşe gelen doğru ve yanlış madde yanıtlarının değerlendirmeye alınması için CUSUM kişi-uyum istatistiği önerilmiştir. Böylece madde alt kümelerindeki sapmaların tespit edilmesi, modele uyumlu olmayan yanıt kalıpları ortaya çıktıkça anlık olarak çıktı veren bir gösterge elde edilmesi amaçlanmıştır (Armstrong ve Shi, 2009; Cizek ve Wollack, 2017; Meijer ve Sijtsma, 2001).

CUSUM belirli bir test katılımcısı için tüm yanıt kalıplarının sapma derecesini temsil etmek üzere tek bir değer hesaplamak yerine, bilgisayar tabanlı bir test sırasında ortaya çıkan anormal modeller hakkında bilgi sağlar. Tıpkı bir kalp grafisinin yaptığı gibi sınav süreci boyunca anormal yanıt kalıpları ortaya çıktıkça tüm çıktı ve sinyalleri izler ve normal sınırlar içerisinde kalan değerlerden sapmaları belirler. Kişi-uyum analizi gerektiren pek çok çalışmada uygulaması olan bir istatistiktir (Bradlow ve diğ., 1998; Cizek ve Wollack, 2017; Montgomery, 2012; van Krimpen Stoop ve Meijer, 2001). CUSUM bir Madde Tepki Kuramı modelinde artık değer tabanlı süreç kontrol yöntemidir ve eşitlik (4) ile ifade edilir (Montgomery, 2012).

$$C_i^+ = \max[0, T_i + C_{i-1}^+]$$

$$C_i^- = \min[0, T_i + C_{i-1}^-] \quad (4)$$

$$Ti = \frac{[x_i - P_i(\theta)]}{\mathrm{n}}$$

T_i madde sayısına göre düzenlenmiş artık değerdir, n test katılımcısının cevapladığı toplam soru sayısıdır ve maddelerin herkes için aynı sayıda olduğu bir test için sabit bir değerdir. C_i^+ ardışık pozitif yanıtlardaki artıkları toplarken büyür, C_i^- ardışık negatif yanıtları toplarken küçülür. Test hilesi bağlamında bir test katılımcısının etiketlenmesi için simülasyon çalışmaları veya diğer yöntemler ile eşik değerleri belirlenir. Üst Sınır (ÜS) ve Alt Sınır (AS) değerleri için C_i^+ > ÜS ve C_i^- < AS durumlarını sağlayan yanıt kalıplarına sahip bireyler hileli test alma davranışı gerçekleştirmiş olarak nitelendirilir (Cizek ve Wollack, 2017; Meijer ve Sijtsma, 2001).

2.1.1.3. ZU3 Kişi-Uyum İstatistiği

Parametrik olmayan ilk birey uyum istatistiği olan G indeksi (Guttman, 1944) türevi olan pek çok kişi-uyum istatistiği geliştirilmiştir (van der Flier, 1977; Tatsuoka ve Tasuoka, 1983). Basit bir ifade ile her bir istatistik için değerlendirme kriteri; testi alan bireylerin kolay maddeleri doğru yanıtlaması, güçlük düzeyi yüksek maddelerde ise yanlış cevaplara gitmesidir. Bu tip cevap örüntüleri Gutmann Mükemmel Tepki Örüntüleri olarak nitelendirilir. Elde edilen değer 0-1 arası bir değer almaktadır ve 1'e yakın değerler uyumlu yanıt kalıplarını, 0'a yakın değerler ise testi alan bireyden beklenmeyen yanıtların göstergesi olarak kabul edilmektedir (Tenderio ve diğ., 2016).

Bir diğer parametrik olmayan U3 kişi-uyum istatistiği de Gutmann Mükemmel Tepki Örüntülerinin tersi olan, kolay maddelerin yanlış ve zor maddelerin doğru olduğu Mokken Monoton Homojenlik modeli bağlamında geliştirilmiştir. Bu istatistik bir test katılımcısnın cevap vektörünün Gutmann ölçeği ile uyumluluğunu denetler ve 0 ile 1 arasında değişen derecede değer alır. Yanıtlar Gutmann ölçeğiyle uyumlu ise 0, sapma gösteriyor ise 1'e yakın değerler alır (van der Flier, 1980; Cizek ve Wollack, 2017).

Bununla birlikte U3 istatistiğinin parametreleri testte yer alan doğru yanıt sayısına bağlıdır. Bu sınırlamayı ortadan kaldırmak adına van der Flier (1982) ZU3'ü önermiştir. ZU3 indeksi U3 indeksinin standartlaştırılmış halidir. ZU3 standart normal dağılıma oldukça yakın bir dağılım göstermekte ve buda yorumlama kolaylığını beraberinde getirmektedir (Emons ve diğ., 2002). Çalışmada kullanılan ZU3 istatistiğinin eşitliği şöyledir (van der Flier, 1982);

$$ZU3 = \frac{\mathrm{U3}-\mathrm{E(U3)}}{\sqrt{Var(U3)}} \tag{5}$$

2.1.2. Kopya ve Benzerlik Ölçüleri

Kopya ve benzerlik ölçüleri test katılımcıları arasındaki eşleşen madde örüntülerini ele alarak farklı yaklaşımlar ile doğru-yanlış eşleşmeleri niceliksel olarak tespit edip anlamlı ölçüde benzerlik gösteren katılımcıları belirlemeyi amaçlar. Her iki yöntem grubu da testi alanların tepki benzerliklerini ölçer, yaklaşımları ve hesaplama prosedürleri farklılık gösterir. en genel tanımı ile bu ölçülerde, istatistiksel model altında tahmin edilen eşleşme sayısı ile ortaya çıkan istatistiğin değeri mukayese edilerek test hilesi bağlamında bir etiketleme gerçekleştirilir. Çoktan seçmeli maddelerden oluşan sınavlarda kopya davranışını belirlemek için literatürde önerilen 20'den fazla yöntem vardır (Sotaridona ve Meijer, 2002, 2003; van der Linden ve Sotaridona, 2006; Wollack, 1996, 2003, 2006; Wollack ve Cohen, 1998; Maynes, 2014).

Literatürde en çok araştırılan ve uygulanan kopya belirleme ve benzerlik indeksleri arasında Genelleştirilmiş Binom Testi (van der Linden ve Sotaridona, 2006), ω istatistiği (Wollack, 1998), K indeksi (Holland, 1996), M4 indeksi (Maynes, 2014), K1 ve K2 indeksleri (Sotaridona ve Meijer, 2002) ve S1 ile S2 istatistikleri (Sotaridona ve Meijer, 2002) bulunmaktadır. Genelleştirilmiş Binom Testi, ω istatistiği ve M4 istatistiği MTK modellerini kullanan istatistiklerdir, K indeksi ve türevleri MTK tabanlı olmayan istatistiklerdir.

Kopya analizinde testi alan iki bireyden bir kaynak ve bir kopyacı tespit edilir ve kopyacının yetenek kestirimi ile kaynağın yanıt örüntüsü arasında gerçekleşen eşleşme hesaplanır. Benzerlik analizinde, kaynak ile kopyacı etiketlemesi gerçekleştirilmeden test katılımcılarının tahmini yetenek seviyeleri göz önüne alınarak modelden beklenen eşleşme sayısı belirlenir. Çalışmada kullanılan K istatistiği kopya analizinde, M4 indeksi ise benzerlik analizinde güçlü sonuçlar veren ve kullanılan programda paket içeriği bulunan ölçüler olduğu

için tercih edilmiştir (Holland, 1996; Sotaridona ve Meijer, 2002).

2.1.2.1. K İstatistiği

K istatistiği, biri kaynak diğeri kopyacı olan iki sınav katılımcısının çoktan seçmeli bir testte yanlış cevaplar arasındaki olağandışı anlaşmanın derecesini değerlendirmek için kullanılabilecek bir istatistiktir. K indeksini hesaplamak için önce kopyacıyla aynı sayıda yanlış yanıtlanmış madde sayısına sahip bir sınav alt grubunu belirlenir ve daha sonra bu gruptaki katılımcıların her biri için eşleşen maddelerin sayısı tespit edilir ardından K indeksi belirlenen değerlere göre hesaplanır ve elde edilen değerin belirli bir anlamlılık düzeyinden düşük değer alması test hilesinin göstergesi olarak kabul edilir (Holland, 1996; Sotaridona ve Meijer, 2002).

Kaynak ve kopyacı olmak üzere iki test katılımcısı arasında araştırılan K istatistiği yalnızca yanlış cevapların eşleşmesine odaklandığından doğru yanıtların uyumu veya boş bırakılan maddeler arasındaki benzerlik değerlendirmeye alınmaz. Yalnızca yanlış yanıtlar arasındaki eşleşme test hilesinin göstergesi olarak kabul edilir ve yanlış cevaplanan maddede kaynak ve kopyacı olan test katılımcısının aynı yanlış cevaba gitmesi cevap kopyalama yoluyla hileli davranışın sağlam bir göstergesidir.

K istatistiği analiz ve yorumlama açısından kullanışlı bir istatistiktir. Bunun yanında kaynak ve kopyacı arasındaki yetenek değişiminden etkilenmesi ve yalnızca yanlış cevaplar üzerinden değerlendirmeye gitmesi açısından sınırlamaları bulunmaktadır. Aynı zamanda kaynak ve kopyacı arasındaki yetenek düzeyi göz önüne alındığında yüksek yetenek düzeyine sahip ikilinin yanlış cevap sayısı ve dolayısıyla eşleşmesi azalacak bu da istatistiğin değerini niceliksel anlamda düşürecektir. Yetenek seviyelerinin düşük olması ise aksi yönde yanlış olarak eşleşen madde sayısını artıracak böylece istatistik değeri artacaktır. Bu farklılaşmayı azaltmak ve daha standart bir değer aralığında sonuç elde etmek için yukarıda bahsedildiği üzere kopyacı bireyin yanlış cevap sayısına göre bir sınav alt grubu oluşturulmaktadır (Sotaridona ve Meijer, 2002; Sotaridona ve Meijer, 2003).

2.1.2.2. M4 İstatistiği

M4 benzerlik istatistiği test katılımcılarının maddelere vermiş oldukları tepkilerin uyuşmasına dayalı bir istatistiktir. Yukarıda bahsedilen K istatistiğinden farklı olarak M4 istatistiği bir benzerlik istatistiğidir ve bireylerin yalnızca yanlış cevaplarına değil aynı zamanda

doğru cevaplarını da değerlendirmeye almaktadır. Test katılımcıları arasında yetenek ve sınav güçlüğü de göz önüne alındığında benzer doğru cevapların gözlemlenmesi, yanlış cevapların gözlemlenmesinden daha olasıdır ve bu yüzden yalnızca doğru cevap eşleşmelerine dayanan bazı kopya-benzerlik istatistiklerine göre M4 istatistiği avantajlı konumdadır. Diğer benzerlik çalışmaları ile karşılaştırıldığında yanıt benzerliğinin olasılığını hesaplamak için en iyi yaklaşımı sapladığı konusunda bir görüş birliği vardır (Cizek ve Wollack, 2017; Maynes, 2014).

M4 istatistiği iki değişkenli bir istatistiktir (aynı doğru cevaplar ve aynı yanlış cevaplar). Bu istatistik ayrık olarak, eşleşen doğru yanıtlar ile eşleşen yanlış yanıtların değerlendirilmesini bütünsel bir biçimde sağladığı için Genelleştirilmiş Binom Testi (van der Linden ve Sotaridona, 2006), K istatistiği gibi diğer olasılık tabanlı kopya-benzerlik araştırmalarından daha avantajlıdır.

M4 istatistiği eşleşmeyen yanıtların olasılığı, eşleşen doğru yanıtların olasılığı, eşleşen yanlış yanıtların olasılığı olmak üzere genelleştirilmiş üç terimli bir dağılım kullanmaktadır. Örtüşme olasılıklarının belirli bir anlamlılık düzeyi üzerinde sonuç vermesi test katılımcılarının cevapları kopyalama, gizli anlaşma gibi test hilesi davranışlarının göstergesi sayılabilir (Maynes, 2014; Maynes, 2017; Zopluoglu, 2019).

2.1.3. Tepki Süresi Modelleri

Test hilesinin belirlenmesi noktasında uzmanlar genellikle kişi-uyum analizlerini önermektedir (Glas ve Dagohoy, 2007; Sinhiray, 2016). Bununla birlikte madde ön bilgisi, hız kontrolü gibi daha indirgenmiş soruşturma çeşitlerinde yine test hilesi belirleme çalışmalarının artarak ilgi gören bir parçası olan madde tepki süresine dayalı istatistiklerin kullanımı tavsiye edilmektedir (Meijer, 2003; Meijer ve Sotaridona, 2006; Steinkamp, 2017). Tepki süresi modelleri testi alanların madde tepki sürelerinin belirli yanıt sürelerine uyumunun değerlendirilmesini içerir. Bazı tepki süresi modelleri tüm test katılımcılarının yanıt süreleri ile örtüşmeyebilir ve söz konusu test katılımcıları çeşitli hile davranışlarını gösterme noktasında hileci alt kümesinde değerlendirilebilir. Bu istatistikler test hilesinin belirlenmesi, sınav katılımcılarının yeteneklerinin kestirilmesi, testin yapısal özelliklerini iyileştirmek ve kişi hızlarını belirlemek için sıklıkla kullanılan yöntemlerdir (Bolsinova ve Tijmstra, 2018; Hanson ve diğ., 1987; Holland, 1996; Marianti ve diğ., 2014; Schnikpe ve diğ., 1999; van der Linden ve Guo, 2008; van der Linden, 2007).

Test hilesinin belirlenmesinde madde tepki süresinin kullanılmasıyla ilgili sınırlı sayıda araştırma bulunmaktadır. Son yıllarda bilgisayar tabanlı sınavların daha yaygın hale gelmesi, beklenmeyen yanıt kalıplarını tespit etmek için madde yanıt sürelerini kullanan yaklaşımlara olan ilgiyi artırmış ve bu konuda gerçekleştirilen çalışmalar ivme kazanmıştır. Bilgisayar tabanlı testler sınava girenlerin tek tek maddeler üzerinde ve sınavın tümünde kullandığı süreler ile ilgili kesin bilgiler sağlar. Bu yüzden bilgisayar tabanlı sınavlarda tepki süreleri bir katılımcının cevap kalıplarının değerlendirilmesinde girdi olarak kullanılabilir. Özellikle madde ön bilgisi başta olmak üzere; madde ön bilgisi, katılımcıların sınav öncesi test maddeleri ve içerik hakkında bilgi sahibi olmasıdır; hileye başvuran katılımcıları ve tehlikeye atılan maddeleri; sınav öncesi belirli bir katılımcısı kümesi tarafından bilgi sahibi olunan maddeler; tespit etmek için tepki süresi modellerinin kullanımının gerçekleştirilen araştırmalarda iyi sonuçlar verdiği ortaya koyulmuştur. Test alma davranışında anormal durumları (tahmin, sınav motivasyonu, rastgele yanıt verme, düşük çaba ve test hilesi) araştırmak için çeşitli tepki süresi modelleri önerilmiştir (Hauser ve diğ., 2011; van der Linden, 2007; Wang ve Xu, 2015; Schnipke ve Scrams, 1999; Wise ve diğ., 2010).

Bunun yanında tıpkı bir kişi-uyum analizi gerçekleştirilmeden evvel MTK modelinin genel uyumunun değerlendirilmesinin gerektiği gibi tepki süresi modelleri için de benzer bir yaklaşımın benimsenmesi gerekmektedir (Meijer ve Tendeiro, 2012; Sinhiray, 2017). Bu yöntemler ile gerçekleştirilen çok sayıda çalışma elde edilen sonuçların bir sınav katılımcısını hileli davranışta bulunduğu noktasında etiketlemede ikincil kanıt olması gerektiği üzerinde durmaktadır (Hanson ve diğ., 1987; van der Linden ve Guo, 2008).

Aşağıda araştırma kapsamında kullanılan iki tepki süresi modelinden kısaca bahsedilmiştir.

2.1.3.1. Log-Normal Tepki Süresi Modeli (LNMRT)

LNMRT (Log-Normal Model for Response Times) test hilesinin belirlenmesi noktasında tepki süresi modelleri altında en sık kullanılan ve güçlü tespit oranları sağlayan modellerin başında gelmektedir. Mevcut model yalnızca yanıt sürelerinin analizinde, madde yanıt sürelerinin madde yanıtları ile ortak analizinde ve madde tepki süreleri ile madde yanıtlarının hiyerarşik olarak modellenmesi gibi pek çok farklı model tasarımı altında kullanılmaktadır ve kendisinden yola çıkarak yapılan bir çok iyileştirme ve geliştirme çalışması

vardır (Bolsinova ve Tijmstra, 2018; Boughton ve diğ., 2017; Glass ve van der Linden, 2010; Qian ve diğ., 2016; Sinharay, 2020; van der Linden ve Guo, 2008; van der Linden, 2007; van der Linden, 2009). Genel model tasarımında N maddeden oluşan bir testte herhangi bir sınav katılımcısının n maddesinde geçirdiği sürenin logaritması alınır;

$$y_n = \log(t_n) \tag{6}$$

ardından ortalaması μ ve standart sapması σ^2 olan N ile ifade edilen log-normal dağılıma yakınsayan bir dağılımda LNMRT eşitliği şu şekilde tanımlanır;

$$y_n|\tau \sim N\left(\beta_n - \tau, \frac{1}{\alpha_n^2}\right) \tag{7}$$

τ değişkeni söz konusu sınav katılımcısının hız parametresini temsil eder. Bu değişkenin daha küçük değerleri testi alan için kendinden beklenen yanıt süresinin üzerinde tepki süreleri ile maddeleri yanıtladığı, daha büyük değerleri ise tam tersi bir biçimde sınav katılımcısından beklenen yanıt sürelerinden daha kısa sürelerde sorulara cevap verdiği anlamına gelmektedir. n maddesi için tüm test katılımcılarının yanıt süreleri göz önüne alınarak kestirilen zaman yoğunluk değişkeni β_n ile ifade edilir. β_n'nin büyük değerleri tüm test katılımcıları için daha büyük beklenen yanıt sürelerine işaret etmektedir. Bir diğer değişken α_n, n. madde için zaman ayrım parametresidir; α_n'nin daha büyük değerleri kişi hızları hakkında daha anlamlı ve fazla bilgiye sahip olunduğu böylelikle hesaplama ve kestirimde daha küçük standart hata olduğu anlamına gelmektedir. LNMRT'nin belirlenebilirliğini sağlamak, incelenen kişi yeteneği hakkında kesitirimde bulunmak ve maksimum olabilirlik yaklaşımı kullanılarak madde parametrelerini tahmin etmek için τ'nin önsel dağılımının LNMRT yaklaşımlarının çoğunda ortalaması 0 ve varyansı σ^2 olan bir normal dağılım olduğu varsayılır (Sinharay, 2020; van der Linden, 2006; van der Linden ve Guo, 2008). Madde parametrelerinin belirlenmesi noktasında Gibbs Örnekleyici, Marjinal Maksimum Olabilirlik Tahmini ve Maksimum Olabilirlik Tahminlerini elde etmek için faktör analizinin kullanılması gibi pek çok yaklaşım araştırmacılar tarafından çalışmalarında kullanılmıştır (van der Linden, 2006; Glas ve van der Linden, 2010).

2.1.3.2. χpf İstatistiği

LNMRT yaklaşımından tepki süresi tabanlı bir istatistik önerilmiştir (Marianti ve diğ., 2014; Fox ve Marianti, 2017). Önerilen istatistik aşağıda ifade edilmiştir;

$$l^t = \sum_n^{\infty} a_n^2(y_n - \beta_n + \tau)^2 \sim \mathcal{X}_N^2 \quad (8)$$

denklemde ifade edilen y_n standartlaştırılmış kalıntıdır ve standart normal dağılımın yapısına uygun bir şekilde dağılır. l^t N serbestlik dereceli ki-kare dağılımını takip etmektedir (Sinharay, 2018; Anderson, 1984). Her bir test katılımcısı için hesaplanan l^t değerleri $\mathcal{X}_N^2$ dağılımının üst %5'lik kesme seviyesinden daha büyük bir değerde ise hileli davranışın göstergesi olabilir ve bu aralıkta kalan bireyler şaibeli katılımcılar kümesinde değerlendirilebilir (Marianti ve diğ., 2014).

Kullanılan ikinci tepki süresi modeli olan χpf istatistiği l^t istatistiğinde hız parametresi olan τ'nin yerine gizli hız parametresi olan $\hat{\tau}$ kullanımıyla farklılık göstermektedir. Bunun yanında χpf standartlaştırılmış kalıntıların kareleri toplamı olarak hesaplanır ve model ile tam uyum gösteren dağılımı N-1 serbestlik dereceli bir ki-kare dağılımını takip eder. Sınava giren herhangi bir test katılımcısı için standart kalıntı şu şekilde hesaplanır;

$$\chi\text{pf} = \sum_n^{\infty} a_n^2(y_n - \beta_n + \hat{\tau})^2 = \sum_n^{\infty} r_n^2 \quad ; \quad \chi\text{pf} \sim \mathcal{X}_{N-1}^2 \quad (9)$$

LNMRT test katılımcısının tepki süreleri ile sapma gösteriyor ise mutlak değer olarak standart kalıntılar büyük değerler olarak çıktı verir ve χpf değeri de büyük olur. Testlerde hile davranışının belirlenmesi için kullanılan istatistiklerde anlamlılık düzeyinin seçimi oldukça önemli bir konudur, tepki süresine dayalı istatistiklerde bunun oldukça makul seviyelerde uygulandığı görülmektedir (Wollack ve diğ., 2015). χpf yanıt sürelerine dayalı frekansçı yaklaşımı benimseyen ilk istatistik olması ve %1'lik anlamlılık düzeyinde dahi oldukça iyi tespit oranları sağlaması ile ümit vaat edicidir (Sinharay, 2018).

2.2. Denetimli Sınıflandırma Yöntemleri

Literatürde denetimli makine öğrenmesinde kullanılan pek çok sınıflandırma tekniği bulunmaktadır. Bunlardan mevcut çalışma için yüksek tespit oranları beklenen üç sınıflandırma tekniği çalışma kapsamında incelenmiştir. Denetimli sınıflandırma yöntemlerinin kullanılması hileli davranışlarda bulunan test katılımcıların belirlenmesinde yeni bir yaklaşım değildir (Man ve diğ., 2019; Zopluoglu 2019). Gerçekleştirilen çalışmada farklı denetimli yöntemler üzerinde durulmuş ve algoritmayı eğitme, girdi verileri gibi hususlarda önceki çalışmalardan ayrı

prosedürler kullanılmıştır. Bu yönüyle daha önceki çalışmalardan daha yüksek tespit oranlarının sağlanabileceği düşünülmektedir.

2.2.1. K En Yakın Komşu Sınıflandırıcısı

K-En Yakın Komşu sınıflandırıcısı denetimli sınıflandırma algoritmaları arasında uzaklığa dayalı olarak sınıflandırma yapan, sınıflandırma ve regresyon problemlerinde sıklıkla kullanılan en etkili ve verimli algoritmaların başında gelmektedir (Gordan ve diğ., 2006). Bunun nedeni bir örneğin bir takım niteliklerinin bilinmesi halinde ona en yakın uzaklıkta olan örneklerin niteliklerinin de belirlenebileceği gibi basit bir mantık çerçevesinde çalışmasıdır. Aynı zamanda parametrik olmayan doğası nedeniyle herhangi bir varsayım gerektirmez ve özellik ve benzerlik yaklaşımına dayalı olduğu için çalıştırılması ve yorumlaması oldukça kolay bir algoritmadır. Yeni etiketsiz gözlemleri algoritmayı tekrar eğitmeye ihtiyaç duymadan ait oldukları sınıfa atayabilir.

Fix ve Hodges (1951) tarafından geliştirilen bu parametrik olmayan kümeleme yaklaşımı daha sonra Cover (1967) genişletilmiş ve yıllar içinde pek çok iyileştirme yapılmıştır (Al-Fahoum ve Al-Fraihat, 2014; Gordan ve diğ., 2006). KNN sınıflandırma algoritması hemen hemen her disiplinde sıklıkla kullanılan çok yönlü bir algoritmadır. Sınıflandırma aşamasında eğitim veri setinin modele önceden sunulmasından dolayı "tembel" veya "hafıza temelli" sınıflandırma olarak da isimlendirilir.

KNN girdi değişkenleri ile çıktı değişkenlerini içeren bir eğitim veri setinden bir tahmin fonksiyonu üretir, ardından bu tahmin fonksiyonuna göre veri setinde bulunan etiketsiz gözlemler sınıflandırılır ve bir gruba atanır. En yakın komşunun belirlenmesinde etiketsiz gözlem ile eğitim kümesindeki örnekler arasında uzaklık ölçümü yapılarak en yakın komşuların çoğunluğunu sağlayan kategoriye yeni örnek atanır. Yani bir birimin sınıflandırılmasında en yakınında yer alan tek bir örneğin dikkate alınmasından ziyade birden çok komşuyu esas alır. K, etiketsiz gözlemin çoğunlukta bulunan en yakın komşularının sayısıdır. Bu değer sınıflandırma öncesinde belirlenir ve değerin yüksek tutulması sınıflandırma doğruluğuna katkı sağlayabilir. K değerinin belirlenmesi noktasında K=1 ile başlanarak sırasıyla değerler sınanır, sınıflandırma doğruluğunda en düşük hata değerini veren K değerine ulaşıncaya kadar bu devam eder. K genellikle tek bir sayıdır ve eğitim veri seti ne kadar büyük ise K değeri de o ölçüde büyük olmaktadır. Etiketsiz gözlemlerin sınıflandırma tahmin fonksiyonuna dahil

edilmesinde mesafe ölçüleri kullanılır. Bu ölçüler Öklid, Hamming, Manhattan ve Minkowski gibi uzaklık ölçüleridir. KNN algoritması etiketsiz gözlemler ile halihazırdaki sınıflandırılmış gözlemler arasındaki benzerliğe odaklanarak yeni gözlemi en çok benzediği gözlemlerin bulunduğu sınıfa atar (Chomboon ve diğ., 2015; Man ve diğ., 2019).

Bu yöntemin başlıca avantajları; uygulamasının ve anlaşılmasının kolay olması, veri seti yeterince büyükse ortalamayı esas aldığı için uç değerlerden az etkilenmesidir. Dezavantajları arasında; seçilen K değeri ile uzaklık ölçüsünden doğrudan etkilenmesi ve eğitim veri setinin çok yüksek tutulması halinde test veri setinde yer alan nesnelerin sınıflandırma etkililiğinin azalması gösterilebilir.

2.2.2. Naive Bayes Sınıflandırıcısı

Olasılık teorisi temelli Bayes teoremi bir olasılık dağılımında herhangi bir rastgele değişken için koşullu olasılıklar ile marjinal olasılıklar arasındaki ilişkiyi gösterir. Hedef değişken ile bağımsız değişkenler arasındaki ilişkiyi analiz eden tahmin edici ve sınıflandırıcı bir algoritmadır. Koşullu bir A olayı için olasılık değeri bağlantılı olduğu bir B olayı için olan olasılık değerinden farklıdır. Yani A olayı gerçekleştiğinde B olayının gerçekleşme olasılığı ile B olayı gerçekleştiğinde A olayının gerçekleşeme olasılıkları farklılık göstermektedir ancak bu iki koşullu olay arasında bir bağlantı bulunmaktadır. İki rastgele olayın koşullu olasılıklarını maksimum olasılık ilkesine dayalı olarak ilişkilendiren ve en yüksek olasılığa sahip sınıfın örneklerini tahmin eden bir algoritmadır. Denetimli sınıflandırma yöntemleri arasında Naive Bayes sınıflandırıcısı yaygın olarak kullanılmaktadır ve sınıflandırma performansı en güçlü olan algoritmaların başını çekmektedir (Ceci, 2005; Challangula, 2005; Menzies ve diğ., 2007; Menzies ve diğ., 2008; Munson ve Khoshgoftaar, 1990; Padberg ve diğ., 2004; Panda ve Patra, 2007; Stigler, 1982; Turhan ve Bener, 2007).

Thomas Bayes (1763) tarafından ortaya koyulan teoremin gelişimine Finette, Savage, Lindley gibi pek çok araştırmacı katkıda bulunmuştur. Tüm iyileştirmelerde olayların olasılıklarının her yeni bilgiyle yeniden ele alınması esas alınmıştır (Jatana ve Sharma, 2014; Lantz, 2013). Bayes teoremi olasılık hesaplamasına dayanır. Bir olay ile ilgili koşulların ön bilgileri göz önüne alınarak o olayın gerçekleşme ihtimalini tanımlar. Diğer sınıflandırma algoritmalarından daha hızlı çalışan bir sınıflandırma tekniğidir. Tüm tahmin edicilerin birbirinden bağımsız olduğu varsayımına dayanır yani bir etiketsiz gözlemin kategorisini

tahmin etmek için her bir girdi değişkeni birbirinden bağımsız olmalıdır.

Teoremin basit matematiksel ifadesinde A ve B ayrı rastgele olaylar olsun;

$$(A|B) = \frac{\mathrm{P(B|A)}P(B)}{\mathrm{P(A)}} \tag{10}$$

P(A), ön olasılık olarak bilinen ve kanıt olarak da adlandırılan A olayının olma olasılığı; P(B), B olayının olma olasılığı; P(A|B), sonsal olasılık olarak da bilinen ve B olayı varken A olayının olma olasılığı; P(B|A), olasılık olasılığı olarak tanımlanan ve A olayı varken B olayının olma olasılığını ifade eder. Bir örneğin hangi sınıfa ait olduğunun belirlenebilmesi için her bir sınıf için olasılıklar hesaplanır ve yeni örnek en yüksek olasılığa sahip sınıfa atanır. NB yaklaşımı bağımlı değişkenlerin birbiri ile ilişkisiz olduğu ve sonuca eşit ölçüde etki ettiği varsayımına dayanır, bu varsayım NB algoritmasının kolay ve verimli çalışmasında en önemli etkendir (Gupta, 2017; Roman, 2018, Tax ve Duin, 2002).

Uygulama kolaylığı ve büyük verilerde kullanışlı ve güçlü doğası ile diğer algoritmalardan daha işlevsel olduğu yönünde yaygın bir görüş birliği vardır. Bir diğer avantajı; çok büyük veri setlerine ihtiyaç duymadan küçük eğitim verileri ile de sınıflama doğruluğu yüksek sonuçlar üretebilmesi ve sürekli veya kesikli veriler ile kullanılabilmesidir. Bunun yanında başlıca dezavantajları; önceki varsayılan tahminlere dayanır ve modele geri dönmeden hatalı tahmin sonuçları üretebilir. Ayrıca önceki sonuçlara güvenilmesinden dolayı bir hata oranı vardır ve girdi değişkenlerinin birbirinden bağımsız olduğu varsayımdan dolayı özellikler arasındaki ilişkiler modellenemez. Son olarak sıfır olasılık probleminin göz ardı edilmesi sorunlara yol açabilir (Gupta, 2017; Roman, 2018; Tax ve Duin, 2002).

2.2.3. Rastgele Orman Sınıflandırıcısı

Rastgele Orman sınıflandırma algoritması Breiman (2001) tarafından geliştirilmiş bir topluluk öğrenme yöntemidir. Temelinde yine Breiman (1996) tarafından geliştirilen Bagging yönteminin ve Ho (1998) tarafından öne sürülen The Random Subspace yöntemlerinin bütünsel görüşü yatmaktadır. Bagging yöntemden farkı değişken seçimi konusunda alt grup oluşturmasıdır. Oluşturulan ağaçlar her bir düğümde önyükleme örnekleri ve toplam tahminci sayısından (N) daha az olan n adet tahminci ile ortaya çıkar ve oluşturulan karar ağacına budama uygulanmaz (N>n) (Breiman, 2001; Cutler ve diğ., 2011; James ve Thomas, 2013).

Bu sınıflandırma yönteminde, {h(x, θk) k=1,…} olarak formüle edilen ağaç tipi bölümlemeli sınıflandırıcılardan faydalanılır (Cutler ve diğ., 2011). Burada x; girdi verilerini, θk; rastgele vektörü temsil eder. Temel veri setinden daha küçük boyutta farklı veri setleri oluşturulur ve bu alt kümelerde rastgele özellik seçimi yöntemiyle yeni ağaçlar yetiştirilir. Bu ağaçların sayısı ile her bir düğümde kullanılacak değişken sayısı kullanıcı tanımlı olarak belirlenir ve ardından değişken sayısı hataları optimum düzeye indirgeyene dek artırılır ya da azaltılır. Değişken sayısının optimum değerden az veya fazla olması güç ve korelasyon üzerinde doğrusal bir etkiye sahiptir. Algoritma yordayıcıları ikili bir şekilde böler, her bir üst düğümden yalnızca iki alt düğüm oluşur. Bu bölünme işleminde Gini indeksinden faydalanılır, sınıf homojenliğinin sağlanması için Gini indeksi minimum değerde tutulur ve sıfıra ulaşınca sınıflama tamamlanır (Breiman, 2001; Pal, 2005; Watts ve diğ., 2011; Archer ve Kimes, 2008; Man ve diğ., 2019).

Bu yöntemin popülerliği ve başlıca avantajı sınıflandırma algoritmaları arasında tahmin geçerliğinin oldukça yüksek olması, modelin yorumlanmasının ve anlaşılması için kolay bir görselleştirme sağlamasıdır. Bunun yanında algoritma durdurma kuralı olarak çeşitli optimizasyon seçenekleri ile zengin bir yapısı vardır ve yüksek boyutlu veri setlerinde verimlidir. Aynı zamanda modele sonradan veri ekleyerek doğruluk oranı yüksek tahminler yapmaya imkan sağlamaktadır (Cutler ve diğ., 2011, Strobl ve diğ., 2009).

BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın modeli, çalışma grubu, veri toplama aracı, veri toplama aracının psikometrik nitelikleri, verilerin toplanması ile verilerin çözümlenmesi ve yorumlanması hakkındaki bilgilere yer verilmiştir.

3.1. Araştırmanın Modeli

Araştırmada, test hilesinin belirlenmesinde kullanılan psikometrik tabanlı Kişi-Uyum İstatistikleri, Kopya ve Benzerlik Ölçüleri ve Tepki Süresi Modelleri ile çeşitli Denetimli Sınıflandırma Algoritmalarının hedeflenen durumu doğru tespit etme oranlarının gerçek bir veri seti ile saptanmasını amaçlamış ve çeşitli model önerileri sunulmuştur. Bu nedenle bu çalışma, farklı yöntemlerin birbiriyle ilişkilerini, farklılıklarını, benzerliklerini ortaya çıkarmayı amaçladığından temel araştırma niteliğindedir.

3.1.1. Önerilen Yaklaşımlar

Bu araştırmada test hilesinin farklı girdi değişkenleri ile farklı modeller kullanılarak daha iyi değerlendirilmesinde ve doğru tespit oranının artırılmasında dört yeni yaklaşım önerilmektedir.

3.1.1.1.Yöntemlerin Birleşik Kullanımı Yaklaşımı

Anormal test davranışlarının tespitinde kanıt olarak tek bir gösterge yerine birden fazla ölçünün kullanılması tavsiye edilmektedir (Meijer, 2003). Testlerde hile davranışlarını değerlendirirken birden fazla ölçüye dayalı kanıt toplanması ve farklı girdi değişkenlerinin değerlendirmeye alınması yalnızca tek bir istatistik ve tek tip girdi değişkeni ile karar vermeye alternatif olan faydalı bir yaklaşım olabilir.

Yöntemlerin birleşik gücünden faydalanmanın amacı, test hilesinin belirlenmesi noktasında çeşitli bakış açılarından kanıt sağlamak ve kritik bir değerlendirme süreci olan test hilesi tespitinin daha sağlam ve güçlü şekilde belirlenmesinin sağlanmasıdır. Birden fazla

göstergeden kanıt toplamanın test hilesinin tespit edilmesinde daha iyi bilgi sağlayacağı düşünülmüştür.

Farklı istatistikler farklı hile davranışlarını tespit etme noktasında daha başarılı ise açımlayıcı bir sınav sonrası adli soruşturma için birden fazla ölçünün birleşik kullanımı daha iyi tespit oranları sağlayabilir. Steinkamp (2017) gerçekleştirdiği araştırmada bunu incelemiş ve daha düşük tip 1 hata oranları tespit etmiştir. Meijer ve Sotaridona (2006) hileli davranışta bulunan bireylerin tespit edilmesinde birden fazla göstergeden yararlanılması gerekliliği üzerinde durmuştur. Meijer ve Sijtsma'da (2001) yaptıkları çalışmada anormal yanıtları etiketlemek için bir Kişi-Uyum Ölçüsü kullanmış ardından anormal yanıtın muhtemel nedenlerini ortaya çıkarmak için etiketlenen bireylerin Kişi Yanıt Eğrilerini incelemişlerdir. Kullanılan bu prosedürler zaman ve hesaplama açısından maliyetli olsa da testlerde hile davranışı hem birey hem de toplumsal açıdan çok ciddi etkileri olan hassas bir durumdur. Birden fazla yöntemin birleşik kullanımı hile davranışına çok yönlü bir bakış açısı sağlayabilir ve doğru tespit oranını iyileştirebilir.

Kişi-Uyum İstatistikleri, Kopya ve Benzerlik Analizleri, Tepki Süresi Modellerinin bazı kombinasyonlar ile kullanılması halinde hileli test alma davranışını daha doğru ve yüksek bir belirleme oranıyla tespit edebileceği öngörülmüştür. Bunun için aralarından yüksek doğruluk oranına sahip istatistikler,

i) Bir Kişi-Uyum İstatistiği ve bir Tepki Süresi Modeli

ii) Bir Kopya ve Benzerlik Ölçüsü ve bir Kişi-Uyum İstatistiği

iii) Bir Tepki Süresi Modeli ve bir Kopya ve Benzerlik Ölçüsü

iv) Bir Kişi-Uyum İstatistiği, bir Kopya ve Benzerlik Ölçüsü ve bir Tepki Süresi Modeli

şeklinde hileli test katılımcılarını içerdikleri alt kümeler birleştirilerek değerlendirmeye alınmıştır.

Önerilen modelin avantajlı sayılabilecek bir diğer yönü; özellikle bilgisayar tabanlı test sistemleriyle eş zamanlı elde edilebilen ikili madde yanıtları, kategorik madde yanıtları, madde yanıt süreleri gibi pek çok veri türünü işleme alarak daha zengin bir analiz fırsatı sağlamasıdır.

Kişi-Uyum İstatistikleri için ikili madde yanıtları, Tepki Süresi Modelleri için madde yanıt süreleri, Kopya ve Benzerlik analizleri için ise kategorik madde yanıtları (A, B, C, D) sırasıyla bu indekslerin hesaplanmasında kullanıldığından daha fazla bilgi ile test hilesinin belirlenmesi sağlanmış olur.

Test katılımcısının sınav davranışını hileli olarak nitelendirmek için tek bir ölçünün kullanılması tavsiye edilmez (Meijer, 2003). Sınava giren kişinin tepki kalıplarının beklenen modele uygunluğuna ilişkin kararların alınmasına yardımcı olmak için birden fazla ölçü kullanılabilir. Bu model şaibeli bir sınav söz konusu olduğunda gözlemsel prosedürler ile incelenecek alt grubu belirlemek için faydalı bir açımlayıcı yaklaşım olabilir. Böylece test güvenliği çalışmalarında oldukça maliyetli ve uzun süreçler gerektiren sınav sonrası izlemeler için tasarruf sağlanabilir. Yapılan çalışmada, her bir Kişi-Uyum İstatistiği, Kopya ve Benzerlik Ölçüsü ve Tepki Süresi Modeli birden fazla kombinasyonda kullanılmıştır.

3.1.1.2. Sınav Süreci Yaklaşımı

Test hilesi eğitsel testler ve değerlendirme alanında giderek daha fazla çalışılan bir konudur. Eğitsel testlerde genel bağlamda 3 tür hileli davranış türü görülmektedir (Wollacak ve Schoening, 2018);

1) Kopya çekme ve gizli anlaşma
2) Madde ön bilgisine sahip olma
3) Sınav sonrası test tahrifatı

Bu hileli davranış türlerini tespit etmek için 5 temel yöntem sınıfı vardır ve pek çoğu Madde Tepki Kuramı tabanlı istatistiklerdir. Bunlar;

1) Puan farklarının incelenmesi
2) Silmelerin ve Cevap Değişikliklerinin Analizi
3) Kopya ve Benzerlik Analizleri
4) Kişi-Uyum İstatistikleri

5) Tepki Süresi Modellerinin kullanılması

Puan farkları madde ön bilgisi ve test tahrifatını, silme ve cevap değişikliklerinin analizi yalnızca test tahrifatını, kopya ve benzerlik analizleri ile kişi-uyum istatistikleri tüm hileli test alma davranış türlerini ve tepki süresi modelleri madde ön bilgisine sahip test katılımcılarını belirlemek için ağırlıkla kullanılır. Aşağıdaki tablo durumu özetlemektedir (Sinharay, 2021);

Tablo 3.1. *Üç Tür Test Hilesi ve Bunları Tespit Etmeye Yönelik Beş Tür İstatistiksel Yaklaşım*

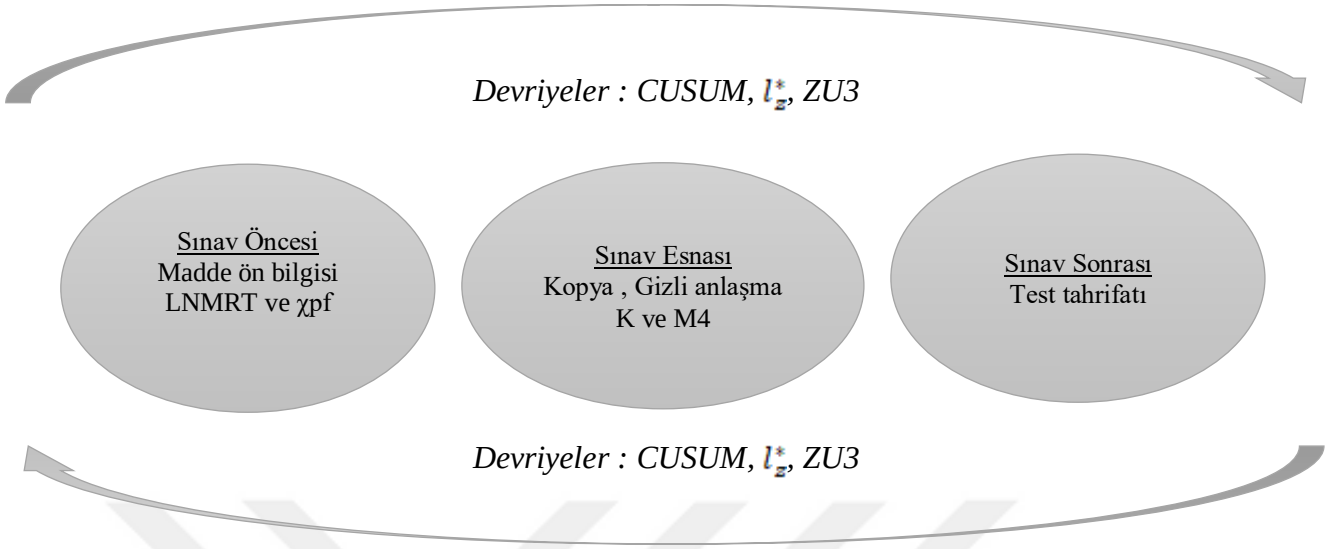
Yaklaşımlar	Cevap Kopyalama ve Gizli Anlaşma	Madde Ön Bilgisi	Test Tahrifatı
Puan Farkı		√	√
Silmelerin ve Cevap Değişikliklerinin Analizi			√
Kopyalama ve Benzerlik Ölçüleri	√	√	√
Kişi-Uyum İstatistikleri	√	√	√
Tepki Süresi Modelleri		√	

Test tahrifatı, en basit haliyle test tamamlandıktan sonra uygulayıcıların veya bizzat bireyin madde yanıtlarında değişikliğe gitmesidir. Bilgisayar uygulamalı testlerde görülmesi nadir bir durumdur daha çok kağıt kalem testlerinde incelenebilecek bir durumdur (Sinharay ve diğ., 2017; Wollack ve diğ., 2015). Veri sağlayıcı tarafından çalışmada kullanılan veri setinde test hilesi davranışlarından cevapları kopyalayan ve madde ön bilgisine sahip katılımcıların olduğu raporlanmıştır. Yukarıda bahsedilen üç tür test hilesi türünden test tahrifatı ile ilgili bildirim yapılmamıştır ve çalışma kapsamında EDI (Wollack ve Eckerly, 2017) veya L-indeksi (Sinharay ve diğ., 2017) gibi cevap değişikliklerinin analizlerini tespit etmeye dönük bir analiz gerçekleştirilmemiştir. Aynı zamanda puan farkı belirleme yaklaşımları da model kapsamına alınmamıştır. Test uygulamasını bir süreç modeli olduğu gerçeği ile test hilesini tespit etmeye yönelik yaklaşımların yalnızca tek tip test hilesi türüne odaklanması amacıyla sınav merkezleri için ayrı ayrı Kopya ve Benzerlik Ölçüleri çalıştırılmıştır. Böylece incelenemeye alınacak merkezlerinde bir listesi elde edilebilir.

Sınav süreci; sınav öncesi, sınav esnası ve sınav sonrası olmak üzere bölümlendiğinde yalnızca sürecin tek bir paydaşına odaklanan ve modelde yalnızca bir hile türünü tespit etme ile sınırlandırılmış istatistikler "bekçiler", süreci bütünsel olarak soruşturan istatistikler ise "devriyeler" olarak nitelendirilmiştir. Farklı hile türlerine bekçilik eden, sınav öncesi madde ön bilgisini tespit etmek için tepki süresi modellerinden LNMRT ile χpf, sınav esnasında cevapları kopyalamayı belirlemek için daha yüksek anlamlılık seviyesinde (ρ=.01) kopya ve benzerlik analizlerinden K ve M4 indeksleri, sınav sonrası ve aynı zamanda tüm sınav süreci boyunca görülebilecek hileli test alma davranışlarını belirlemek için kişi-uyum istatistiklerinden l_z^*, CUSUM ve ZU3 (devriyeler) .01 kesme seviyesinde çeşitli birleşimler ile değerlendirmeye alınmıştır. Sınav Süreci Yaklaşımı Tablo 3.2. ve Şekil 3.1. ile özetlenmektedir;

Tablo 3.2. *Önerilen Sınav Süreci Yaklaşımına Göre Test Hilesi Türleri ve Bunları Tespit Etmeye Yönelik İstatistiksel Yaklaşımlar*

Yaklaşımlar	Madde Ön Bilgisi	Cevap Kopyalama ve Gizli Anlaşma	Test Tahrifatı
Kişi-Uyum İstatistikleri	√	√	√
Kopya ve Benzerlik Ölçüleri		√	
Tepki Süresi Modelleri	√		



Şekil 3.1. Önerilen Sınav Süreci Yaklaşımı

Önerilen sınav süreci yaklaşımı için Kişi-Uyum İstatistikleri, Kopya ve Benzerlik Ölçüleri ile Tepki Süresi Modellerinden dört farklı alt model değerlendirmeye alınmıştır. İlk model devriyeler ile birlikte LNMRT ve K indeksini, ikinci model devriyeler ile birlikte χpf istatistiği ve K indeksini, üçüncü model devriyeler ile LNMRT ve M4 indeksini, son model ise devriyeler ile birlikte χpf istatistiği ve M4 indeksini kapsamaktadır.

3.1.1.3.Denetimli Sınıflandırma Yöntemlerinde Basit Topluluk Oyu Yaklaşımı

Denetimli sınıflandırma yöntemleri çıktıları içinde yöntemlerin bütünsel gücünden yararlanılabilir. Basit Topluluk Oyu olarak isimlendirilen modelde her bir yöntemin bir test katılımcısını hileci veya normal olarak etiketlemesi üzerinden bir çoğunluk oylamasına gidilir. Çalışma koşullarında birden fazla sınıflandırma yönteminin hileli test alma davranışı saptadığı bireyler etiketlenir. Tablo 3.3., çalışmadan bir kesit sunulmaktadır.

Tablo 3.3. *Denetimli Sınıflandırma Yöntemlerinde Test Veri Seti Çıktıları ve Etiketli Katılımcılar İçin Basit Topluluk Oyu Yöntemi (1 : Test Hilesi Yapan Bireyler ve 0: Normal Sınav Katılımcıları*)

Katılımcı Numarası	KNN	Naive Bayes	Rastgele Ormanlar	Topluluk Oyu Sayısı
2	1	1	0	2
3	1	1	0	2
6	0	0	0	0
8	1	1	1	3
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:
1620	1	1	1	3
1622	0	1	0	0
1631	0	0	0	0

3.1.1.4. Denetimli Sınıflandırma Yöntemlerinde Birleşim Yaklaşımı

Önerilen ilk yaklaşıma benzer şekilde farklı denetimli sınıflandırma yöntemlerinin hileli test alma davranışını belirlemede oluşturdukları alt kümelerin birleşimine gidilerek birden fazla algoritmanın çıktıları bir araya getirilebilir. Bu modeller basit ve kullanışlı çözümler getirebilir. Tablo 3.4., çalışmadan bir kesit sunulmaktadır.

Tablo 3.4. *Denetimli sınıflandırma yöntemlerinde test veri seti çıktıları ve etiketli katılımcılar için birleşim yaklaşımı (1 : Test Hilesi Yapan Bireyler ve 0: Normal Sınav Katılımcıları)*

Katılımcı Numarası	KNN	Naive Bayes	Rastgele Ormanlar	Birleşim Yaklaşımı
2	1	1	0	+
3	1	1	0	+
6	0	0	0	-
8	1	1	1	+
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:
:	:	:	:	:
1620	1	1	1	+
1622	0	1	0	+
1631	0	0	0	-

Birden fazla göstergenin kullanılmasında her bir psikometrik tabanlı istatistikle elde edilen potansiyel hileci test katılımcılarının alt kümesinin birleşimine ardından elde edilen birleşim kümesinin veri sağlayıcı tarafından hileli davrandığı yönünde bildirim yapılan 46 katılımcının kümesi ile kesişimi gerçekleştirilmiştir.

Önerilen modellerden sınav süreci yaklaşımı bağlamında istatistiksel analiz sürecinde hile yapılan merkezlerin belirlenmesi ve test hilesi türlerinden yalnızca cevap kopyalama davranışına odaklanmak amacıyla sınav merkezleri bazında bir analiz gerçekleştirilmiştir. Sözdiziminde “centers” argümanı kullanılarak bütün merkezlerde olası tüm çiftler arasında M4 ve K indeksleri .01 anlamlılık düzeyinde değerlendirmeye alınmıştır. Kişi-Uyum İstatistikleri için de kesme seviyeleri .01 değerinde tutulmuş ve yanlış pozitif oranının düşürülmesi amaçlanmıştır.

3.2. Araştırmanın Materyali

Farklı yöntemlerin duyarlılık, özgüllük ve doğruluk oranlarını araştırılmasında Cizek ve Wollack'dan (2017) alınan gerçek bir veri kümesi kullanılmıştır. Veri setine R Programı LNIRT paket içeriğinden ulaşılabilir (Fox ve diğ., 2021). Veri kümesi her sene düzenli olarak uygulanan bilgisayar tabanlı bir testten gelmektedir. Uygulanan test MTK puanlamasını ve testi ölçeklemek için Rasch Modelini kullanmaktadır. Veri setinin hileli test alma davranışında bulunan bireyleri içerdiği bilindiğinden test güvenliği çalışmalarında hile tespit yöntemlerini incelemek için ideal olduğu düşünülmüştür. Test sağlayıcısı tarafından rapor edilen 46 katılımcı hileli davranış sergileyen katılımcılar olarak etiketlenmiştir. Bu katılımcıların bir kısmının madde ön bilgisi, bir kısmının ise cevapları kopyalama yolu ile test hilesi yaptığı bildirilmiştir. Etiketlenen test katılımcılarının tümünün test hilesi davranışı gösterdiği, test hilesinde bulunan etiketlenmesi gereken ancak tespit edilemeyen başka katılımcıların da olabileceği Cizek ve Wollack (2017) tarafından çalışmalarında vurgulanmıştır.

Veri setinde her bir katılımcının 170 madde için madde yanıt bilgileri bulunmaktadır. Veri seti her bir katılımcı için 10 demografik giriş değişkenini, 170 ikili puanlanmış madde yanıtı (doğru yanıtlar için 1, yanlış yanıtlar için 0), 170 dörtlü derecelendirilmiş madde yanıtını (A, B, C, D şıkları), madde tepki sürelerini (her bir madde için saniye cinsinden), toplam süreyi içermektedir. Toplamda 1636 test katılımcısının verisi bulunmaktadır.

3.3. Veri Analiz Aracı

Analizler eğitsel değerlendirmeler ve sezgisel yöntemlerde oldukça sık kullanılan ve açık kaynak kodlu R Programı ile gerçekleştirilmiştir. Kişi-Uyum İstatistiklerinin analizi R Programı (versiyon 4.1.2) ile, Kopya ve Benzerlik Analizleri R Programı (4.1.2)'nin kullanılan paketi desteklememesinden dolayı Microsoft R Open Programı (4.0.2) ile, denetimli sınıflandırma yöntemleri R Studio Programı ile gerçekleştirilmiştir. Çıktılar dosyalanmış ve Notepad++ Programı ile sözdizimi kaydedilmiştir (R Core Team, 2018).

3.4. Veri Analiz Süreci

Araştırmada hileli test alma davranışı sergileyen bireyleri belirlemek için psikometrik tabanlı Kişi-Uyum İstatistikleri, Kopya Ve Benzerlik Ölçüleri ve Tepki Süresi Modelleri;

psikometrik tabanlı yöntemlerden önerilen iki ayrı yöntem; Denetimli Sınıflandırma Yöntemleri ve Denetimli Sınıflandırma Yöntemlerinden önerilen iki bütünsel bakış açısı kullanılmıştır.

3.4.1. Kişi-Uyum İstatistiklerinin Test Hilesinin Belirlenmesinde Kullanımı

Araştırma bağlamında Kişi-Uyum İstatistikleri için ikili yanıtlar değerlendirmeye alınmıştır. l_z^* ve ZU3 analizleri R Programı 'PerFit' paketi (Tendeiro ve diğ., 2016) ile gerçekleştirildi. CUSUM Kişi-Uyum İstatistiğinin hesaplanmasında R Programı 'PsyControl' paketi (Hong ve Can, 2017) kullanıldı. Her bir istatistikte madde parametrelerini tahmin etmede İki Parametreli Lojistik Model (2PLM) kullanılmıştır. Yine kişi yeteneklerinin tahmininde Maksimum Olabilirlik yöntemi tercih edilmiştir. Herhangi bir MTK modelinin öngördüğü yetenek düzeyleri göz önüne alınarak hileli test alma davranışlarını belirlemede kesme değerleri kullanılmıştır. Bunlar hangi değerden sonra test katılımcılarından elde edilen yanıt kalıplarının modelden beklenenden uyumsuz olarak sınıflandırılması gerektiğinin ölçütleridir. Kesme değerlerinin belirlenmesinde alanyazında net bir görüş birliği yoktur. Bazı araştırmacılar hileli davranışı belirlemek için sabit değerler önermiştir (Huang, 2012; Snijders, 2001; Karabatsos, 2003; Sijtsma ve Meijer, 1992). Fakat kesme puanları madde sayısı, testi alanların yetenek düzeyleri, sınavın zorluğu gibi pek çok değişkenden etkilenmektedir ve tüm testlerde standart olarak kabul gören kesme değerleri hatalı sonuçlar doğurabilir. Bu değişkenleri göz önüne alarak farklı Kişi-Uyum İstatistikleri, Kopya ve Benzerlik Ölçüleri ve Tepki Süresi Modelleri için kesme puanları deneysel dağılımlar ile belirlenir. Her istatistik için kesme puanları farklıdır. Testlerde hile yapan katılımcıların tespit edilmesine yönelik istatistiklerde kesme puanlarını belirlemede kullanılan yaklaşımlar; sabit bir değeri ölçüt olarak almak, gerçek veri örneğinden yola çıkarak gerçekleştirilen simülasyon çalışmalarından elde edilen dağılımların belirli bir alt-üst yüzdelik dilimini kullanmak, mevcut test verilerinden yeniden örnekleme yoluyla kesme değerlerinin belirlenmesi, ρ değeri kullanılarak koşullu bir sınıflandırma yapmak ve ilk etapta anormal tepki kalıpları olan bireyleri veri setinden çıkararak sonrasında elde edilen madde ve kişi parametreleri ile indekslerin hesaplanarak kesme puanlarının belirlenmesidir (Mousavi ve diğ., 2019; Patton ve diğ., 2019).

Kesme noktaları l_z^* ve ZU3 istatistikleri için .05, CUSUM için üst kuyruk .975 ve alt kuyruk .025 olarak belirlendi. Bu kesme noktaları kullanılan programda varsayılan ve

literatürde sıklıkla kullanılan değerler olduğu için seçilmiştir (Man ve diğ., 2019; Steinkamp, 2017). Modele uygun oluşturulan örnekleme dağılımlarında kesme puanlarını belirlemek için l_z^* ve ZU3 istatistiklerinde 1000 yeniden örnekleme ve her birinde 10000 varsayımsal test katılımcısını içeren bir prosedür izlenmiştir. CUSUM istatistiği için belirlenen önyükleme örneği sayısı 1000 olarak belirlenmiştir. Böylece hesaplama maliyetini kısmen yükselterek daha yüksek tespit oranı sağlayan sonuçlar elde edilmesi amaçlanmıştır.

3.4.2. Kopya Ve Benzerlik Ölçülerinin Test Hilesinin Belirlenmesinde Kullanımı

Kopya ve Benzerlik Ölçüleri olan M4 ile K istatistiklerinin hesaplanmasında kategorik madde yanıtları analize alınmıştır. Her iki istatistiğin analizi R Programı 'CopyDetect' paketi (Zopluoglu ve Zopluoglu, 2018) ile gerçekleştirilmiştir. Gerçekleştirilen prosedürde yöntemlerin birleşik kullanımındaki göstergeler için ρ=.05 istatistiksel anlamlılık düzeyinde tüm çiftler arasında hesaplanan M4 ve K İndeksleri değerlendirmeye alınmıştır. Wollack ve Schoening'e (2018) göre Kopya ve Benzerlik Analizleri madde ön bilgisini, cevap kopyalamayı, gizli anlaşmaları ve test tahrifatını belirlemek için kullanılabilir. Her iki yaklaşımda da M4 İndeksi için Rasch Modeli kullanılmıştır.

3.4.3. Tepki Süresi Modellerinin Test Hilesinin Belirlenmesinde Kullanımı

Log-Normal Tepki Süresi Modellemesi (LNMRT) için analizler R Programı 'LNIRT' paketi (Fox ve diğ., 2021) kullanılarak gerçekleştirilmiştir. Daha yüksek duyarlılık ve özgüllük oranlarınının sağlanması amacıyla MCMC yineleme sayısı 5000, artıklar hesaplanmadan gerçekleştirilecek MCMC yinelemelerinin sayısı ise 1000 olarak girilmiştir.

Sinhiray (2018) tarafından tepki süreleri için log-normal bir model olarak önerilen χpf analizi için oluşturulan R kod dizilimi kullanılmıştır. Verilerin analizinde kesme değerinin hesaplanmasında bir başka R Paketi olan 'PerFit' paketindeki (Tendeiro ve diğ., 2016) "cutoff" fonksiyonundan yararlanılmıştır. Elde edilen kesme puanından daha yüksek bir eşik değeri olan "205" değeri yapılan farklı denemeler sonucunda daha optimal bir kesme puanı olarak belirlenmiştir. Bu kesme puanı test hilesi yapan katılımcıların alt kümesinin belirlenmesinde yanlış pozitif oranını düşürmek ve tespit oranını iyileştirmek için tercih edilmiştir.

3.4.4. Denetimli Sınıflandırma Yöntemlerinin Test Hilesinin Belirlenmesinde Kullanımı

Testlerde hile davranışı gerçekleştiren bireylerin daha etkili ve verimli bir biçimde tespit edilebilmesi için bir test sürecinde elde edilen tüm bilgilerden nasıl yararlanılacağı konusunda yaygın bir görüş birliği ve yeterli çalışma yoktur. Makine öğrenmesi yöntemleri bu sorunu ele almak için iyi çözümler sunabilir. Testi alanlar ile ilgili tüm değişkenler girdi değişkenleri olarak ortaklaşa modellenebilir ve tüm test sahtekarlığı türlerini kapsayan bir yaklaşımla sınav sonrası istatistiksel soruşturma için makine öğrenmesi yöntemleri kullanılabilir (Man ve diğ., 2019).

Psikometrik tabanlı test hilesi belirleme yaklaşımlarının sınırlı olduğu durumlar vardır. Özellikle bilgisayar tabanlı testlerden elde edilen geniş hacimli verileri birleştirmek ve işlemek için yeterince donanımlı değildirler. Aynı zamanda yalnızca hile davranışının tüm alt boyutları ile soruşturulması için kullanışlı olmayabilirler. Denetimli Sınıflandırma Algoritmaları tüm test hilesi türlerini bütün girdi değişkenlerini değerlendirmeye alarak tespit edebilir.

Çalışmada Denetimli Sınıflandırma Algoritmalarından KNN Sınıflandırıcısı, Naive Bayes Sınıflandırıcısı, Rastgele Ormanlar Sınıflandırıcısı kullanılmıştır. Denetimli Sınıflandırma Yöntemleri için sınıflar iyi bir biçimde eğitilmiş modellere dayanmaktadır. Genel olarak önce bir model oluşturulacak daha sonra oluşturulan modele göre yeni bir birim sınıflandırılacaktır. Veri setindeki katılımcıların sınıf etiketine göre %70'i Denetimli Sınıflandırma Algoritmalarını eğitmek, kalan %30'luk kısım model doğrulaması ve tespit oranının belirlenmesi için kullanılmıştır.

3.4.4.1.Girdi Verilerinin Oluşturulması

Denetimli Sınıflandırma Yöntemleri için yalnızca test yönetimi sırasında elde edilen bilgiler değil aynı zamanda psikometrik yaklaşımlara ait indeksler, türetilmiş değişkenler de algoritmaya dahil edilerek test hilesini belirlemede daha yüksek tespit oranlarına ulaşılması hedeflenmiştir.

Kişi yetenekleri R Programı 'irtoys' paketi (Partchev ve diğ., 2017) kullanılarak kestirilmiştir. Kişi hızları için R Programı 'LNIRT' paketinden (Fox ve diğ., 2021) zaman parametreleri kestirilmiştir. Ardından 'lattice' ve 'ltm' paketleri kullanılarak çalışma kapsamında oluşturulan kod dizilimi ile kişi hız parametreleri kestirilmiştir (Sarkar ve diğ.,

2015; Rizopoulos ve Rizopoulos, 2018).

Denetimli sınıflandırma yöntemleri için değerlendirmeye alınan girdi veri seti;

a) madde yanıtları (ikili),

b) madde yanıt süreleri,

c) toplam süre,

d) toplam doğru yanıtlardan oluşan ve 0 ile 170 arasında değişen genel başarı puanı,

e) kişi hızları,

f) kişi yetenekleri,

g) l_z^* Kişi-Uyum İstatistiği,

h) ZU3 Kişi-Uyum İstatistiği,

i) CUSUM Kişi-Uyum İstatistiği pozitif artıklar toplamı,

j) CUSUM Kişi-Uyum İstatistiği negatif artıklar toplamı,

k) χpf puanları,

l) testi alanların hileli davranışlarına yönelik etiket sınıfları

değişkenlerini içermektedir. Toplamda 350 farklı değişken analize dahil edilmiştir.

3.4.4.2. Özellik Ölçekleme

Denetimli Sınıflandırma Yöntemleri için bir diğer husus özellik ölçekleme olarak da adlandırılan girdi değişkenlerini ortak bir ölçeğe aktarma işlemi olan veri normalizasyonudur. Verileri sınıflandırmaya hazır hale getirmek için tüm verilerin aynı aralığa getirilmesi veri ön işleme sürecinde oldukça önemli bir adımdır. Birim uzunluğa göre ölçekleme, varyansa göre ölçekleme, ortalama bazlı ölçekleme, TanH, maksimum ve minimum değerlere göre ölçekleme ve T-skoru gibi verileri normalleştirmede kullanılan pek çok yöntem vardır (Fukunaga, 2013;

Juszczak ve diğ., 2002; Man ve diğ., 2019). Mevcut çalışmada girdi verileri minimum ve maksimum değerlere göre ölçeklendirilmiştir. Bu tip bir özellik ölçeklemede veriler 0 ile 1 arasında değerlere sahiptir ve daha kolay karşılaştırılabilir sonuçlar sağlar. Bunun yanında ikili madde yanıtlarını ölçekleyebilmesi ve hesaplamasının kolay olması (x-min(x))/(max(x)-min(x)) pratik açıdan araştırma verileri için uygundur. Bunun için R Programında oluşturulan fonksiyon kullanılmıştır.

3.4.4.3. Özellik Seçimi

Modelden çıkarımlar yapmak için verilerin normalleştirilmesinden sonra girdi değişkenleri arasından bağımsız değişkenler olarak sınıflamada kullanılacak değişkenler seçilecektir. Özellik seçiminin uygulanması daha az karmaşık bir modelin oluşturulması, modelden doğru çıkarımların yapılması ve hesaplama maliyeti açısından verimli olabilecek bir yaklaşımdır. Bu çalışmada özellik seçimi için Özyinelemeli Özellik Eliminasyonu (Recursive Feature Elimination) yöntemi kullanılmıştır. Özyinelemeli Özellik Eliminasyonunun temel amacı sınıf etiketlerine göre bağımsız değişkenlerin önem katsayılarını belirlemek ve en önemli tahmin değişkenlerinden başlayarak modele en çok katkı sağlayan değişkenleri tespit etmektir. Değişkenlerin seçimi özyinelemeli olarak yapılır, bunun gerekçesi bazı değişkenlerin öneminin farklı alt kümelere göre değişiklik gösterebilmesidir (Granitto ve diğ., 2006). Özyinelemeli Özellik Eliminasyonu modele en az katkı sağlayan özellikleri aşamalı olarak ortadan kaldırarak girdi veri setini en iyi temsil eden optimum alt kümeyi belirlenmesini sağlar.

Bunun için R Programı ‘caret’ paketi (Kuhn ve diğ., 2020) ve harici yeniden örnekleme yöntemi olarak 10-kat çapraz doğrulama kullanılmıştır.

Gerçekleştirilen özellik seçimi işlemi sonrası denetimli sınıflandırma yöntemlerine sunulan son girdi seti; kişi hızlarını, toplam süreyi, genel başarı puanını, kişi yeteneklerini, ZU3 kişi-uyum istatistiğini, ve 13 farklı maddeye ait madde tepki süresi verilerini içermektedir.

Girdi verilerinin oluşturulması, özellik ölçekleme ve özellik seçiminin ardından Denetimli Sınıflandırma Yöntemleriyle test hilesinin belirlenmesinde K-En Yakın Komşu Algoritması, Naive Bayes Sınıflandırma Algoritması ve Rastgele Orman Sınıflandırma Algoritmalarının analizleri gereçekleştirilmiştir. Denetimli Sınıflandırma Yöntemlerinden K-En Yakın Komşu Algoritmasının çalıştırılmasında R Programı ‘class’ paketi (Ripley ve diğ.,

2015) kullanılmıştır. Bu paket uygulama yönergelerinin açık ve kolay olduğu, KNN yöntemi için sıklıkla kullanılan ve iyi kabul edilen bir paket olduğu için tercih edilmiştir. Naive Bayes Sınıflandırma Algoritmasının uygulanması için R Programı 'e1071' paketi kullanılmıştır (Meyer ve diğ., 2019). Rastgele Ormanlar Sınıflandırma Algoritması için R Programı 'randomForest' paketi (RColorBrewer ve Liaw, 2018) tercih edilmiştir. Her bir girdi değişkeni için oluşturulacak ağaç sayısı 1000 olarak belirlenmiştir. Kalan parametrelerde varsayılan değerler geçerli kalmıştır.

3.4.5. Sonuç Ölçütleri

Bir yöntemin doğru tespit yeterliliğini incelemek için kullanılan çeşitli ölçütler vardır. Gerçekleştirilen çalışmada hileli test alma davranışı gerçekleştiren katılımcıların yöntem tabanlı sınıflandırmalarında sonuç ölçütleri olarak Duyarlılık (Sensitivity), Özgüllük (Specificity) ve Doğruluk (Accuracy) ölçüleri kullanılmıştır. Duyarlılık, test sağlayıcı tarafından sınav yönetimi sırasında hileli test alma davranışı gerçekleştiği rapor edilen ve herhangi bir yöntemle hileli davranışı saptanan katılımcıların yüzdesidir. Özgüllük, test yöneticisi tarafından hileye başvurmayan normal test katılımcıları olarak nitelendirilip bir yöntem ile normal olarak sınıflandırılan bireylerin yüzdesidir. Doğruluk ise, Duyarlılık ve Özgüllüğün birleştirilerek tek bir ölçü şeklinde ele alınmasıdır. Yani hileli test alma davranışı sergileyenler ve hileli test alma davranışında bulunmayan bireyler olarak toplam doğru tespit oranına "Doğruluk (Accuary) denir. Daha açıklayıcı olması açısından aşağıdaki tablo düzenlenmiş ve ölçütlerin hesaplanışı özetlenmiştir.

Tablo 3.5. *Sonuçları Değerlendirirken Karşılaşılabilecek Durumlar*

	Gerçek Durum		
Yöntem Sonucu	**Hileci Birey**	**Normal Birey**	**Toplam**
Hileci Birey	A	B	A + B
Normal Birey	C	D	C + D
Toplam	A + C	B + D	A + B + C + D

Tabloda;

A : Test sağlayıcı tarafından hileli davran katılımcılar olarak etiketlenen ve belirli bir yöntem ile hile yaptığı tespit edilen olgulardır (Doğru Pozitif).

B : Normalde hileli davranmayan fakat bir yöntem tarafından hileli davranışlar sergilediği sonucuna ulaşılan olgulardır (Yanlış Pozitif).

C : Gerçek durumda test hilesine başvuran fakat değerlendirme sonrası normal davrandığı kabul edilen olgulardır (Yanlış Negatif).

D : Test yöneticisi tarafından normal olarak etiketlenen ve bir yöntem sonucu hile yapmadığı yönünde karar verilen olgulardır (Doğru Negatif).

Ölçütlerin hesaplanışı;

Duyarlılık = A / (A + C)

Özgüllük = D / (B + D)

Doğruluk = (A + D) / (A + B + C + D)

şeklinde özetlenebilir.

Söz konusu ölçütlerin hesaplanması R Programında oluşturulan kodlar ile gerçekleştirilmiştir.

3.4.6. Önerilen Yöntemlerin Analizi

3.4.6.1. Yöntemlerin Birleşik Kullanımı İle Test Hilesinin Belirlenmesi

İki ve üç farklı psikometrik tabanlı yaklaşımın birlikte kullanılmasında Kişi-Uyum İstatistikleri, Tepki Süresi Modelleri ve Kopya ve Benzerlik Ölçülerinden birer ölçü kullanılmıştır. Her bir ölçünün tespit ettiği hileli davranan sınav katılımcılarının alt kümelerinin birleşimine gidilmiştir. l_z^* ve ZU3 analizinde ikili madde yanıtları değerlendirmeye alınarak R Programı 'PerFit' (Tendiro ve diğ., 2016) paketinden yararlanılmış ve kesme seviyesi olarak .05 değeri kullanılmış, CUSUM içinde benzer şekilde ikili madde yanıtları değerlendirmeye

alınarak R Programı 'PsyControl' (Hong ve Can, 2017) paketinden faydalanılmış ve kesme seviyesi üst kuyruk .975 ile alt kuyruk .025 değerleri esas alınmıştır. Her Kişi-Uyum İstatistiği için parametre tahmininde 2PL modeli, yetenek kesitirimi için Maksimum Olabilirlik Tahmini ve 1000 yeniden örnekleme kullanılmıştır. Kopya ve Benzerlik Ölçülerinde M4 ve K İstatistikleri için kategorik madde yanıtları değerlendirmeye alınmış ρ=.05 anlamlılık düzeyinde bir analiz gerçekleştirilmiştir. Her iki Kopya ve Benzerlik Ölçüsünün analizi R Programı 'CopyDetect' paketi ile gerçekleştirilmiştir (Zopluoglu ve Zopluoglu, 2018). Tepki Süresi Modellerinde sınav katılımcılarının maddelere verdikleri yanıt süreleri saniye cinsinden değerlendirmeye alınmış ve LNMRT için MCMC yineleme sayısı 5000 olarak girilmiş, χpf için kesme puanı 205 olarak belirlenmiştir. LNMRT modellemesi için R Programı 'LNIRT' paketi (Fox ve diğ., 2021) ve χpf analizi için R Programında oluşturulan kodlar kullanılmıştır. Ardından sırasıyla "union" komutuyla l_z^* ve l^t; ZU3 ve K; CUSUM ve M4; ZU3 ve l^t; l^t ve K; M4 ve χpf; ZU3 ve LNMRT ile K; CUSUM ve χpf ile K; χpf ve K ile l_z^*; ZU3 ve χpf ile K; CUSUM ve LNMRT ile M4 analizlerinden elde edilen hileci alt kümlerinin birleşimine gidilmiş ardından "intersect" komutu ile veri sağlayıcı tarafından hileli davrandığı yönünde bildirim yapılan 46 katılımcının kümesi ile kesişimi gerçekleştirilmi duyarlılık, özgüllük ve genel doğruluk değerleri hesaplanmıştır.

3.4.6.2. Sınav Süreci Yaklaşımı İle Test Hilesinin Belirlenmesi

Önerilen modellerden sınav süreci yaklaşımı bağlamında istatistiksel analiz sürecinde hile yapılan merkezlerin belirlenmesi ve test hilesi türlerinden yalnızca cevap kopyalama davranışına odaklanmak amacıyla sınav merkezleri bazında bir analiz gerçekleştirilmiştir. Devriyeler olarak nitelendirilen Kişi-Uyum İstatistikleri l_z^*, ZU3 ve CUSUM analizinde ikili madde yanıtları değerlendirmeye alınmış, R Programı 'PerFit' (Tendiro ve diğ., 2016) ve 'PsyControl' (Hong ve Can, 2017) paketleri kullanılmıştır. l_z^* ve ZU3 İstatistikleri için kesme seviyesi olarak .01 değeri, CUSUM için üst kuyruk .995 ile alt kuyruk .005 değerleri kullanılarak yanlış pozitif oranının düşürülmesi amaçlanmıştır. Parametre tahmininlerinde 2PL modeli, yetenek kesitirimleri için ise Maksimum Olabilirlik Tahmini ve 1000 yeniden örnekleme kullanılmıştır. Sınav öncesi madde ön bilgisinin tespitinde bekçilik eden LNMRT istatistiği için, R Programı 'LNIRT' paketinden (Fox ve diğ., 2021) yararlanılmış ve 5000 MCMC yineleme sayısı ile artıklar için 10000 tekrar değeri kullanılmış, χpf istatistiğinin analizi için 205 kesme seviyesi ile kendiliğinden oluşturulan kod ile gerçekleştirilmiştir. Sınav esnası

kopya ve gizli anlaşmaların tespitinde tercih edilen göstergeler M4 ve K istatistikleri için R Programı 'CopyDetect' paketi (Zopluoglu ve Zopluoglu, 2018) sözdiziminde "centers" argümanı kullanılarak bütün merkezlerde olası tüm çiftler arasında M4 ve K İndeksleri ρ=.01 anlamlılık düzeyinde değerlendirmeye alınmıştır. M4 İstatistiğinin analizlerinde yine Rasch Modeli tercih edilmiştir. Sınav sonrası test tahrifatı için veri sağlayıcısı tarafından bir dönüt sunulmamıştır ve devriyeler olası bir sınav sonrası hile davranışı için duyarlı durumdadır. Bu yüzden bu paydaş için bekçilik işlevi görecek istatistikler modele ayrıca dahil edilmemiştir. Ayrı gerçekleştirilen analizlerin ardından son olarak "union" komutuyla sırasıyla, devriyeler (l_z^*, CUSUM ve ZU3), LNMRT ile K; devriyeler, χpf ile K; devriyeler, LNMRT ile M4 ve devriyeler, χpf ile M4 analizlerinden elde edilen hileci alt kümlerinin birleşimine gidilmiştir. Son olarak doğru tespit oranının sınanması için "intersect" komutu ile veri sağlayıcı tarafından hileli davrandığı yönünde bildirim yapılan 46 katılımcının kümesi ile kesişim kümesi oluşturularak duyarlılık, özgüllük ve genel doğruluk değerleri hesaplanmıştır.

3.4.5.3 Denetimli Sınıflandırma Yöntemlerinde Basit Topluluk Oyu Ve Birleşim Yaklaşımı İle Test Hilesinin Belirlenmesi

Burada da benzer şekilde veri setimizin %70'i Denetimli Sınıflandırma Algoritmalarını eğitmek için kalan %30'luk kısım model doğrulaması ve tespit oranının belirlenmesi için kullanılmıştır. Girdi veri setini; ikili madde yanıtları, madde yanıt süreleri, toplam süre, genel başarı puanı, kişi hızları, kişi yetenekleri, l_z^* Kişi-Uyum İstatistiği, ZU3 Kişi-Uyum İstatistiği, CUSUM Kişi-Uyum İstatistiği pozitif artıklar toplamı, CUSUM Kişi-Uyum İstatistiği negatif artıklar toplamı, χpf skorları, testi alanların hileli davranışlarına yönelik etiket sınıfları olmak üzere 350 farklı değişken oluşturmaktadır. Veri sağlayıcı tarafından sunulan değişkenlere ilave olarak kişi yetenekleri R Programı 'irtoys' paketi, kişi hızları için R Programı 'LNIRT' paketi, kişi hız parametreleri R Programı 'lattice' ve 'ltm' paketleri kullanılarak hesaplanmıştır (Partchev ve diğ., 2017; Rizopoulos ve Rizopoulos, 2018; Sarkar diğ., 2015; Fox ve diğ., 2021). Değişken değerleri minimum ve maksimum değerlere göre ölçeklendirilmiş, yeniden örnekleme yöntemi olarak 10-kat çapraz doğrulama kullanılarak R Programı 'caret' paketi (Kuhn ve diğ., 2020) ile son girdi veri seti oluşturulmuştur. Ardından K-En Yakın Komşu Algoritması için 'class' paketi, Naive Bayes Sınıflandırma Algoritması için 'e1071' paketi ve Rastgele Ormanlar Sınıflandırma Algoritması için 'randomForest' paketi kullanılarak varsayılan değerlerde analizler R Programında gerçekleştirişmiştir (Ripley ve diğ., 2015; Meyer

ve diğ., 2019; RColorBrewer ve Liaw, 2018). Son olarak R Programında oluşturulan kodlar ile Basit Topluluk Oyu Yaklaşımı için denetimli sınıflandırma yöntemlerinde iki veya daha fazla yöntemin hileci olarak etiketlediği katılımcılardan bir alt küme oluşturulmuş ve veri sağlayıcı tarafından bildirilen küme ile kesişimine gidilerek raporlanmıştır. Birleşim yaklaşımında ise her bir denetimli sınıflandırma yönteminin etiketlediği katılımcılar tek bir küme altında toplanmış ve veri sağlayıcı tarafından hileli davrandığı yönünde bildirim yapılan 46 katılımcının kümesi ile kesişim kümesi oluşturularak duyarlılık, özgüllük ve genel doğruluk değerleri hesaplanmıştır.

BÖLÜM IV

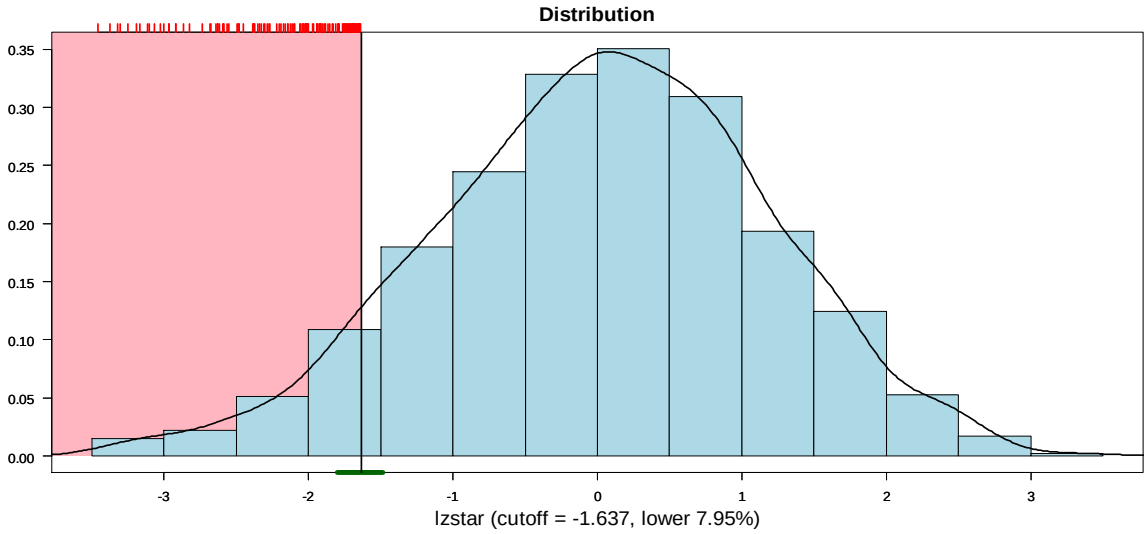
BULGULAR

Bu bölümde araştırmanın amacı, problemi ve alt problemleri doğrultusunda R Programında gerçekleştirilen Kişi-Uyum İstatistikleri olan l_z^* , ZU3 ve CUSUM istatistiklerine, Kopya ve Benzerlik ölçüleri olan K ve M4 indekslerine, Tepki Süresi Modellerinden LNMRT ve χpf göstergelerine, yöntemlerin birleşik kullanımına, önerilen süreç modelindeki alt modellere, denetimli sınıflandırma yöntemlerinden KNN Sınıflandırıcısı, NB Sınıflandırıcısı, RO Sınıflandırıcısına ve denetimli öğrenme yöntemlerinden yola çıkarak önerilen Basit Topluluk Oyu ve Birleşim Yaklaşımına ait bulgulara yer verilmiştir. Sırasıyla alt problemlere ilişkin sonuç ölçütleri olan özgüllük, duyarlılık ve doğruluk çıktıları sunulmuştur.

4.1. Kişi-Uyum İstatistikleri İle Test Hilesinin Belirlenmesine İlişkin Bulgular

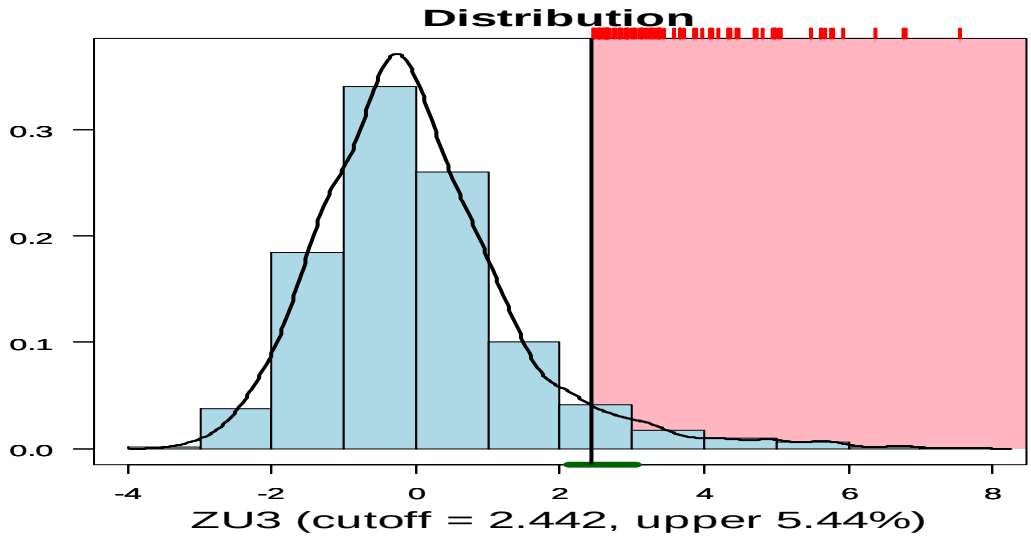
Bu bölümde psikometrik tabanlı Kişi-Uyum İstatistiklerinin analizler sonucunda elde edilen değerlerine ve dağılımlarına yer verilmiştir. l_z^*, ZU3 ve CUSUM Kişi-Uyum İstatistiklerinin sonuç ölçütleri olan özgüllük, duyarlılık ve doğruluk oranları sunulmuştur.

l_z^* Kişi-Uyum İstatistiğinin kesme değeri -1.637 olarak belirlenmiştir. Test grubunun %7.95'i bu eşik değerinin altında kalmış ve testlerde hile davranışı gösteren bireyler olarak etiketlenmiştir (130 kişi). Etiketlenen 130 katılımcı arasında test sağlayıcısı tarafından hile yaptığı onaylanan 4 katılımcı vardır. Yani l_z^* Kişi-Uyum İstatistiği test sağlayıcısı tarafından hileli davranış gösterdiği tespit edilen bireylerin %9'unu başarıyla belirlemiştir. Aynı zamanda hile yapmayan test katılımcılarının doğru tespit oranı %93'tür. Bu normal davranan test katılımcılarının bir kısmının (126 kişi) hileli davranan alt grupta değerlendirildiği anlamına gelmektedir. İstatistiğin dağılımı Şekil 4.1.'de görülmektedir.



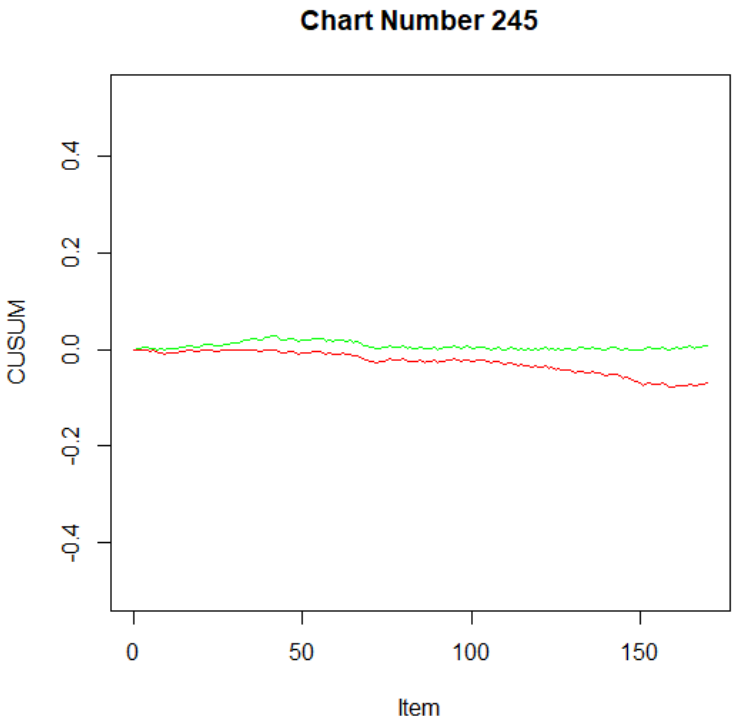
Şekil 4.1. l_z^* İstatistiğinin Dağılımı, Kesme Değeri ve Hileli Davranışta Bulunan Alt Grubun Seçilmesi

ZU3 Kişi-Uyum İstatistiğinin kesme değeri 2.442 olarak belirlenmiştir. Test grubunun %5.44'ü bu eşik değerinin üstünde kalmış ve testlerde hile davranışı gösteren bireyler olarak etiketlenmiştir (89 kişi). Bu bireylerden 5'i test şirketi tarafından rapor edilen katılımcılardır. İstatistiğin hileli davranış gösteren bireyleri %11 başarıyla tespit ettiği sonucuna ulaşılmıştır. Yine hileli davranış göstermeyen test katılımcılarının güvenli bir şekilde belirlenmesindeki başarı oranı %95'tir. 84 birey test sağlayıcısı tarafından normal davranan grupta değerlendirildiği halde hileci olarak etiketlenmiştir. Her birey daha sonra değerlendirmeye alınmak üzere dosyalanmıştır. İstatistiğin ve bireylerin istatistik dağılımı Şekil 4.2.'de görülmektedir.

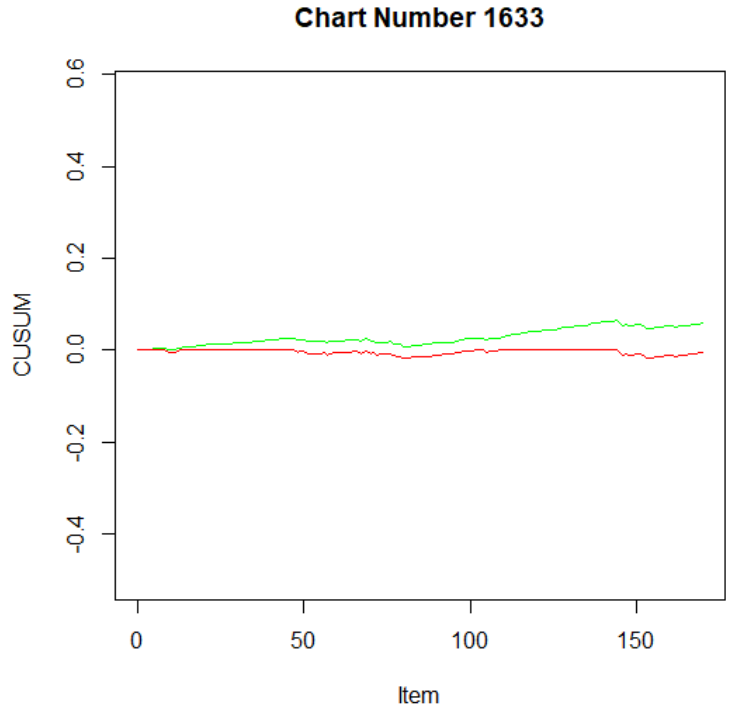


Şekil 4.2. ZU3 İstatistiğinin Dağılımı, Kesme Değeri ve Hileli Davranışta Bulunan Alt Grubun Seçilmesi

CUSUM Kişi-Uyum İstatistiği için alt sınır değer -0.07880413, üst sınır değer 0.07141019 olarak belirlenmiştir. Çalışmada 54 test katılımcısından 35'inin pozitif ardışık artıklarının toplamı 0.07141019 eşiğinin üzerinde ve 19'unun negatif ardışık artıklarının toplamı 0.07141019 eşiğinin altında kalarak test hilesi göstermiş olabilecek alt grup olarak belirlenmiştir. Test sağlayıcısı tarafından hileli davranıdığı yönünde etiketlenen 3 katılımcı bu 54 kişilik alt grupta yer almaktadır. Yani gösterge hileli davranış gösteren bireylerin %7'sini başarılı bir şekilde belirlemiştir. Korunan normal test katılımcılarının oranı %97'dir. Negatif ardışık artıkları toplamı kesme seviyesinin altında olan sınav katılımcısı 245 ve pozitif ardışık artıkları toplamı kesme seviyesinin üstünde olan sınav katılımcısı 1633 için CUSUM süreç çizelgeleri aşağıda sunulmuştur.



Şekil 4.3. Sınav Katılımcısı 245 İçin CUSUM Süreç Çizelgesi



Şekil 4.4. Sınav Katılımcısı 1633 İçin CUSUM Süreç Çizelgesi

Kişi-Uyum İstatistiklerinin duyarlılık, özgüllük ve doğruluk ölçüleri Tablo 4.1.'de sunulmaktadır.

Tablo 4.1. *Kişi-Uyum İstatistiklerinde Duyarlılık, Özgüllük ve Doğruluk*

	Duyarlılık	**Özgüllük**	**Doğruluk**
$\boldsymbol{l_z^*}$	%9	%93	%90
ZU3	%11	%95	%93
CUSUM	%7	%97	%94

l_z^* Kişi-Uyum İstatistiği 129 test katılımcısını hileli test alma davranışı gösteren birey olarak etiketlemiştir. Bunun yanında bu alt grupta gerçek hileci 4 test katılımcısı bulunmaktadır. İstatistiğin duyarlılık oranı %9, özgüllük %93 ve doğruluk oranı %90'dır.

ZU3 Kişi-Uyum İstatistiği 81 test katılımcısını hileli test alma davranışı gösteren birey olarak etiketlemiştir. Bunun yanında bu alt grupta gerçek hileci 5 test katılımcısı bulunmaktadır. İstatistiğin duyarlılık oranı %11, özgüllük %95 ve doğruluk oranı %93'tür.

CUSUM Kişi-Uyum İstatistiği 54 test katılımcısını hileli test alma davranışı gösteren birey olarak etiketlemiştir. Bunun yanında bu alt grupta gerçek hileci 3 test katılımcısı bulunmaktadır. İstatistiğin duyarlılık oranı %7, özgüllük %97 ve doğruluk oranı %94'tür.

Tablo 4.5.'e bakıldığında Kişi-Uyum İstatistiklerinin duyarlılık, özgüllük ve doğruluk oranlarının birbirine yakın değerler aldığı ve doğru tespit oranlarının birbirine yakın olduğu görülmektedir. Her üç Kişi-Uyum İstatistiğinin de duyarlılık, özgüllük ve doğruluk oranları birbirine yakın ve düşük seviyededir.

4.2. Kopya Ve Benzerlik Ölçüleri İle Test Hilesinin Belirlenmesine İlişkin Bulgular

Bu bölümde psikometrik tabanlı kopya ve benzerlik ölçülerinin analizler sonucunda elde edilen değerlerine ve dağılımlarına yer verilmiştir. K ve M4 Kopya ve Benzerlik ölçüleri için sonuç ölçütleri olan duyarlılık, özgüllük ve doğruluk oranları Tablo 4.2.'de sunulmuştur.

Tablo 4.2. *Kopya ve Benzerlik Ölçülerinde Duyarlılık, Özgüllük ve Doğruluk*

	Duyarlılık	**Özgüllük**	**Doğruluk**
M4	%46	%77	%76
K	%20	%93	%91

Tablo 4.2. incelendiğinde, M4 İndeksi 387 test katılımcısını hileli test alma davranışı gösteren birey olarak etiketlediği görülmektedir. Bunun yanında bu alt grupta gerçekten hile davranışı sergileyen 21 test katılımcısı bulunmaktadır. İstatistiğin duyarlılık oranı %46'dır. Bu test sağlayıcısı tarafından bildirilen hileci bireylerin tek bir gösterge ile en yüksek seviyede doğru tespit edildiği istatistik olduğu anlamına gelmektedir. Öte yandan özgüllük (%77) ölçüsü normal davranan test katılımcılarının daha büyük bir kısmının hileli davranışta bulunan katılımcılar olarak etiketlendiğini göstermektedir. K İndeksi 124 test katılımcısını hileli test alma davranışı gösteren birey olarak etiketlemiştir. Bunun yanında bu alt grupta gerçekten hile davranışı sergileyen 9 test katılımcısı bulunmaktadır. İstatistiğin duyarlılık oranı %20, özgüllük %93 ve doğruluk oranı %91'dir. M4 İndeksinin K İndeksine göre oldukça iyi bir oranda test hilesine başvuran bireyleri belirleyebileceği tespit edilmiştir. Bu bulgu, hile yapmayan ve yapmadığı teyit edilen bireyleri belirlemede K İndeksinin daha iyi çalıştığının göstergesidir.

4.3. Tepki Süresi Modelleri İle Test Hilesinin Belirlenmesine İlişkin Bulgular

Bu bölümde psikometrik tabanlı yaklaşımlardan Tepki Süresi Modellerinin analizleri sonucu elde edilen sonuçlara ve dağılımlara yer verilmiştir. LNMRT ve χpf ölçüleri için sonuç ölçütleri olan duyarlılık, özgüllük ve doğruluk oranları Tablo 4.3.'te sunulmuştur.

Tablo 4.3. *Tepki Süresi Modellerinde Duyarlılık, Özgüllük ve Doğruluk*

	Duyarlılık	Özgüllük	Doğruluk
LNMRT	%35	%87	%86
χpf	%33	%90	%88

Literatürde sıklıkla görüldüğü üzere LNMRT Tepki Süresi Modeli ile χpf Tepki Süresi Modelinin sonuç ölçütleri birbirine oldukça yakındır. LNMRT 16 bireyi, χpf 15 bireyi doğru şekilde test hilecisi olarak belirlemiştir. İstatistiklerin sırasıyla duyarlılık oranları %35 ve %33, özgüllükleri %87 ve %90, doğruluk oranları %86 ve %88'dir.

4.4. Denetimli Sınıflandırma Yöntemleri İle Test Hilesinin Belirlenmesine İlişkin Bulgular

Bu bölümde Denetimli Sınıflandırma Algoritmalarına ilişkin analizler sonucunda elde edilen duyarlılık, özgüllük ve doğruluk oranlarına yer verilmiştir. Denetimli Sınıflandırma Algoritmaları olan K-En Yakın Komşu, Naive Bayes ve Rastgele Ormanlar Algoritmalarının duyarlılık, özgüllük ve doğruluk ölçüleri içerisinde veri sağlayıcı tarafından hileli davranış gösterdiği raporlanan 14 test katılımcısı bulunan toplamda 491 sınav katılımcısından elde edilmiştir. Bu ölçüler Tablo 4.4.'te sunulmaktadır.

Tablo 4.4. *Denetimli Sınıflandırma Yöntemleri İçin Duyarlılık, Özgüllük ve Doğruluk*

	Duyarlılık	Özgüllük	Doğruluk
K-En Yakın Komşu	%43	%99	%99
Naive Bayes	%64	%99	%99
Rastgele Ormanlar	%57	%99	%99

Tablo 4.4. incelendiğinde, Denetimli Sınıflandırma Yöntemlerinden olan K-En Yakın Komşu, Naive Bayes ve Rastgele Ormanlar Sınıflandırma Algoritmalarının oldukça yüksek özgüllük ve doğruluk oranları verdiği görülmektedir (%99). Testlerde hile davranışını sergileyen katılımcıları belirleme noktasında en iyi tespit oranını %64 duyarlılık oranıyla Naive Bayes yöntemi gerçekleştirmiştir. Rastgele Ormanlar Sınıflandırma Algoritması ve K-En Yakın

Komşu Algoritmalarının duyarlılık oranları sırasıyla %57 ve %43'tür. Tüm denetimli sınıflandırma yöntemleri için test hilesi yapan katılımcıları belirleme oranı normal davranan katılımcıları hatalı sınıflandırma yapmaksızın diğer yöntemlerden belirgin düzeyde yüksektir.

4.5. Çalışma Kapsamında Önerilen Yaklaşımlara İlişkin Bulgular

4.5.1. Yöntemlerin Birleşik Kullanımı İle Test Hilesinin Belirlenmesine İlişkin Bulgular

Bu bölümde psikometrik tabanlı yaklaşımların birleşik kullanımına ilişkin analizler sonucunda elde edilen duyarlılık, özgüllük ve doğruluk oranlarına yer verilmiştir.

Bir Kişi-Uyum İstatistiği ve bir Tepki Süresi Modeli, bir Kopya ve Benzerlik Ölçüsü ve bir Kişi-Uyum İstatistiği ile bir Tepki Süresi Modeli ve bir Kopya ve Benzerlik Ölçüsünün birlikte kullanımlarının bazı kombinasyonlarına ilişkin duyarlılık, özgüllük ve doğruluk oranları Tablo 4.5.'te sunulmuştur.

Tablo 4.5. *İki Yöntemin Birleşik Kullanımı İçin Duyarlılık, Özgüllük ve Doğruluk*

	Duyarlılık	**Özgüllük**	**Doğruluk**
$\boldsymbol{l_z^*}$ **+ LNMRT**	%41	%83	%83
ZU3 + K	%30	%88	%86
CUSUM + M4	%50	%76	%75
ZU3 + LNMRT	%41	%83	%82
LNMRT + K	%41	%83	%83
M4 + χpf	%63	%70	%68

Tablo 4.5. incelendiğinde, bir Kişi-Uyum İstatistiği ve bir Tepki Süresi Modeli, bir Kopya Ve Benzerlik Ölçüsü ve bir Kişi-Uyum İstatistiği ile bir Tepki Süresi Modeli ve bir Kopya Ve Benzerlik Ölçüsünün birlikte kullanımlarının test hilesinin belirlenmesinde değişen derecelerde doğru tespit oranları sunduğu görülebilir. En yüksek duyarlılık oranı M4 ve χpf ölçülerinin kullanımında gözlemlenmiştir (%63), bunun yanında en düşük özgüllük (%70) ve doğruluk (%68) oranlarına da bu iki kopya-benzerlik ve tepki süresi modelinin birlikte kullanımında rastlanmaktadır. Gerçekleştirilen kombinasyonların ortalama bir çıktısı olarak

CUSUM ve M4 Ölçülerinin duyarlılık, özgüllük ve doğruluk oranları sırasıyla %50, %76 ve %75'tir. En düşük duyarlılık oranını ZU3 ve K İndekslerinin birlikte kullanımı vermiştir (%30) ve geri kalan özgüllük ve doğruluk oranları sırasıyla %88 ve %86 olarak incelemeye alınan birlikte modeller altında en yüksek özgüllük ve duyarlılık oranıdır. Aynı duyarlılık oranına sahip (%41), l_z^* Kişi-Uyum İstatistiği ve LNMRT Tepki Süresi Ölçüsü, ZU3 Kişi-Uyum İstatistiği ve LNMRT Tepki Süresi Ölçüsü ile K Kopya ve Benzerlik İndeksi ve LNMRT Tepki Süresi Ölçülerinin birlikte kullanımları tamamen aynı özgüllük oranını göstermiş (%83), nispeten aynı doğruluk oranları (sırasıyla %83, %82, %83) ortaya çıkarmışlardır. Genel olarak hileli davranış gösteren bireylerin tespit oranları yükseldikçe normal davranan test katılımcılarının da hileli davranan grupta hatalı etiketlenme oranı artmaktadır.

Bir Kişi-Uyum İstatistiği, bir Kopya ve Benzerlik Ölçüsü ve bir Tepki Süresi Modelinin birlikte kullanımlarının bazı kombinasyonlarına ilişkin özgüllük, duyarlılık ve doğruluk oranları Tablo 4.6.'da sunulmuştur.

Tablo 4.6. *Üç Yöntemin Birleşik Kullanımı İçin Duyarlılık, Özgüllük ve Doğruluk*

	Duyarlılık	**Özgüllük**	**Doğruluk**
ZU3 + LNMRT + K	%50	%77	%76
CUSUM + χpf + K	%43	%78	%77
χpf + K + l_z^*	%50	%74	%75
ZU3 + χpf + K	%70	%64	%65
CUSUM + LNMRT + M4	%65	%68	%67

Bir Kişi-Uyum İstatistiği, bir Kopya ve Benzerlik Ölçüsü ve bir Tepki Süresi Modelinin birlikte kullanımların test hilesinin belirlenmesinde değişen derecelerde doğru tespit oranları sunduğu görülebilir. En yüksek duyarlılık oranı ZU3, χpf ve K Ölçülerinin kullanımında gözlemlenmiştir (%70), bunun yanında en düşük özgüllük (%64) ve doğruluk (%65) oranlarına da bu üç göstergenin birlikte kullanımında rastlanmaktadır. Gerçekleştirilen birleşimlerin ortalama bir çıktısı olarak ZU3, LNMRT ve K Ölçülerinin duyarlılık, özgüllük ve doğruluk oranları sırasıyla %50, %77 ve %76'dır. En düşük duyarlılık oranını CUSUM, χpf ve K indekslerinin birlikte kullanımı vermiştir (%43) ve geri kalan özgüllük ve doğruluk oranları

sırasıyla %78 ve %77 olarak incelemeye alınan birlikte modeller altında en yüksek özgüllük ve duyarlılık oranıdır. Aynı duyarlılık oranına sahip (%50) ZU3 Kişi-Uyum istatistiği, LNMRT Tepki Süresi Modeli, K Kopya ve Benzerlik Ölçüsü ile χpf Tepki Süresi Modeli, K Kopya ve Benzerlik Ölçüsü ve l_z^* Kişi-Uyum İstatistiği birleşimlerinin birbirine çok yakın özgüllük (%77-%74) ve doğruluk (%74-%75) gösterdiği belirlenmiştir. Yöntemlerin birlikte kullanımında hile yaptıkları düşünülen alt grubun sayısı arttıkça hileli test alma davranışı gerçekleştiren bireyleri tespit etme oranları artmakta, bunun bir yan ürünü olarak hileli davranışlarda bulunmayan bireyleri hileci olarak nitelendirme oranları da yükselmektedir. Duyarlılık oranının artması özgüllük ve doğruluk oranlarında sürekli olarak düşüşe yol açmıştır. Özellikle dördüncü satıra bakıldığında özgüllük oranının duyarlılığın altında kaldığı tek sonucu ZU3, χpf ve K göstergelerinin birlikte kullanılmasının verdiği görülmektedir.

Yöntemlerin birlikte kullanımında hile yaptıkları düşünülen alt grubun sayısı arttıkça hileli test alma davranışı gerçekleştiren bireyleri tespit etme oranları artmakta, bunun bir yan ürünü olarak hileli davranışlarda bulunmayan bireyleri hileci olarak nitelendirme oranları da yükselmektedir. Duyarlılık oranının artması özgüllük ve doğruluk oranlarında sürekli olarak düşüşe yol açmıştır.

4.5.2. Önerilen Sınav Süreci Modeline İle Test Hilesinin Belirlenmesine İlişkin Bulgular

Bu bölümde önerilen süreç modeline ilişkin analizler sonucunda elde edilen özgüllük, duyarlılık ve doğruluk oranlarına yer verilmiştir.

CUSUM ve ZU3 Kişi-Uyum İstatistikleri, Kopya ve Benzerlik Ölçülerinden K ve M4 İndeksleri ve LNMRT ile χpf Tepki Süresi Modellerinden önerilen dört farklı süreç modeline ilişkin özgüllük, duyarlılık ve doğruluk oranları Tablo 4.7.'de sunulmuştur.

Tablo 4.7. *Sınav Süreci Yaklaşımında Farklı Modeller İçin Duyarlılık, Özgüllük ve Doğruluk*

	Duyarlılık	**Özgüllük**	**Doğruluk**
Model 1	%48	%88	%87
Model 2	%46	%85	%84
Model 3	%50	%83	%82
Model 4	%48	%80	%79

(**Model 1:** Devriyeler, LNMRT ve K, **Model 2:** Devriyeler, χpf ve K, **Model 3:** Devriyeler, LNMRT ve M4, **Model 4:** Devriyeler, χpf ve M4)

Tablo 4.7. incelendiğinde, önerilen modellerin birbirine yakın duyarlılık, özgüllük ve doğruluk oranları verdiği görülmektedir. ZU3 ve CUSUM Kişi-Uyum İstatistikleri ile birlikte (devriyeler), LNMRT Tepki Süresi Ölçüsü ve M4 Kopya ve Benzerlik İndeksi en yüksek duyarlılık oranını vermektedir (%50), bunun yanında özgüllük ve doğruluk oranları da kendinden daha düşük duyarlılık oranı veren kalan üç modele yakındır (sırasıyla %83 ve %82). Aynı duyarlılık oranına sahip (%48) Model 1 ve Model 4, özgüllük ve doğruluk ölçüleri açısından farklılaşmaktadır. Model 1 %88 özgüllük ve %87 doğruluk oranı ile Model 4'ün %80 özgülük ve %79 doğruluk ölçüsünden daha yüksek bir tespit oranı sağlamıştır. Model 1'de devriyelere ek olarak LNMRT Tepki Süresi Modeli ve K Kopya ve Benzerlik Ölçüsü; Model 4'te χpf Tepki Süresi Modeli ve M4 Kopya ve Benzerlik ölçüsü modele alınmıştır.

Bunun yanında merkezler baz alınarak bir değerlendirme gerçekleştirilmesi sonucunda veri sağlayıcı tarafından toplamda kopya çekildiği bildirilen 81 sınav merkezinden 14'ü mevcut analiz ile tespit edilmiştir. %17 doğru tespit oranına tekabül eden bu değer kopya çekilen merkezlerin belirlendiği yaklaşımlar göz önüne alındığında ortalamanın üstünde bir tespit oranı sağlamaktadır (Maynes, 2014; Sinharay, 2018).

4.5.3. Denetimli Sınıflandırma Yöntemlerinde Basit Topluluk Oyu Ve Birleşim Yaklaşımı İle Test Hilesinin Belirlenmesine İlişkin Bulgular

Bu bölümde Denetimli Sınıflandırma Algoritmalarından yola çıkarak önerilen yaklaşımlara ilişkin analizler sonucunda elde edilen özgüllük, duyarlılık ve doğruluk oranlarına

yer verilmiştir.

Denetimli Sınıflandırma Algoritmaları olan K-En Yakın Komşu, Naive Bayes ve Rastgele Ormanlar Algoritmalarının katılımcılar bazında test veri çıktıları ile önerilen Basit Topluluk Oyu Yöntemi ve önerilen Birleşim Yaklaşımı ile elde edilen sonuçlara ait kesitsel bir gösterim Tablo 4.8.'de sunulmaktadır.

Tablo 4.8. *Denetimli Sınıflandırma Yöntemlerinde Basit Topluluk Oyu ve Birleşim Yaklaşımı (1 : Test Hilesi Yapan Bireyler ve 0: Normal Sınav Katılımcıları*)

Katılımcı Numarası	KNN	Naive Bayes	Rastgele Ormanlar	Basit Topluluk Oyu Yaklaşımı	Birleşim Yaklaşımı
2	1	1	0	2	+
3	1	1	0	2	+
6	0	0	0	0	-
8	1	1	1	3	+
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
1620	1	1	1	3	+
1622	0	1	0	0	+
1631	0	0	0	0	-

Denetimli Sınıflandırma Algoritmaları olan K-En Yakın Komşu, Naive Bayes ve Rastgele Ormanlar Algoritmalarının önerilen Basit Topluluk Oyu Yöntemi ve önerilen Birleşim Yaklaşımı ile elde edilen özgüllük, duyarlılık ve doğruluk oranları Tablo 4.9.'da sunulmuştur.

Tablo 4.9. *Denetimli Sınıflandırma Yöntemlerinden Önerilen Yaklaşımlar İçin Duyarlılık, Özgüllük ve Doğruluk*

	Duyarlılık	**Özgüllük**	**Doğruluk**
Basit Topluluk Oyu Yaklaşımı	%67	%97	%96
Birleşim Yaklaşımı	%68	%94	%94

Tablo 4.9. incelendiğinde Denetimli Sınıflandırma Yöntemlerinden önerilen yaklaşımların test hilesi yapan katılımcıları doğru tespit etme oranları oldukça yüksektir. Basit Topluluk Oyu Yöntemi için duyarlılık %67, Birleşim Yaklaşımı için %68'dir. Aynı zamanda sırasıyla özgüllük oranları %97 ve %94'tür. Bu her iki yaklaşımında hileci bireyleri normal olarak davranan bireylerden iyi bir şekilde ayırarak yüksek bir tespit gücü ile belirleyebildiği anlamına gelmektedir.

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde araştırma nihayetinde ulaşılan çıkarımlara ve bu çıkarımlar neticesinde uygulayıcılar ve sonraki araştırmacılar için bir takım önerilere yer verilmiştir.

5.1. Sonuç ve Tartışma

Testlerde hileli davranışların tespit edilmesi konusunda birbirinden farklı pek çok yöntem ve istatistik geliştirilmiştir ve günümüzde devam eden pek çok çalışma bulunmaktadır. Test şirketleri, araştırmacılar ve test uygulayıcılarının ortak ilgi alanı olan bu durum özellikle geniş kapsamlı ve önemli değerlendirmelerin gerçekleştirildiği sınavlar için hayati bir önem taşır. Hileli test alma davranışında bulunan katılımcıların puanlarının değerlendirmeye dahil edilmesinin testin güvenliğini, geçerliğini ve test kullanımının tutarlılığını zayıflattığı bilinmektedir. Test güvenliği araştırmalarında genellikle tek bir yöntem sınıfına, girdi olarak yalnızca tek bir değişken türüne veya hile davranışının tek bir boyutuna odaklanıldığı görülmektedir. Test hilesinin belirlenmesinde özellikle kritik ve yüksek risk düzeyine sahip koşullar göz önüne alındığında optimum tespit gücünün nasıl sağlanacağı sorusu henüz etkili bir şekilde aydınlatılmamıştır.

Halihazırda hileli test alma davranışlarının tespitinde ilgili çalışmaların salt çoğunluğunun istatistiklerin türetilmesi ve geliştirilmesi ile anormal yanıt kalıplarını belirleme oranına odaklandığını belirtilmektedir (Karabatsos, 2003; Meijer, 2003; Meijer ve Sijtsma, 2001). Araştırmada test hilesinin tespitinde kullanılan Kişi-Uyum İstatistikleri ile Kopya ve Benzerlik Ölçülerinin duyarlılık ve özgüllük oranları literatürdeki çalışmalardan nispeten daha yüksek düzeyde gözlemlenmiştir (Man ve diğ., 2019; van der Linden, 2009; van der Linden ve Guo, 2008). Tepki Süresi Modellerinden LNMRT test hilesinin tespitinde önceki çalışmalara yakın bir duyarlılık ve özgüllük değerine sahiptir (Man ve diğ., 2019; Wollack ve diğ., 2015). Bunun yanıda χpf istatistiği önceki çalışmalardan daha yüksek sayılabilecek bir duyarlılık ve özgüllük değerindedir (Sinharay, 2020; Sinharay ve diğ., 2017). χpf istatistiğinin alandaki

çalışmalara göre daha iyi bir tespit oranı sağlamasının temel nedeni istatistik hesaplanırken kesme puanını belirleme yönteminin farklı bir mantık ile yürütülmesi ve normalde ortaya çıkan kesme değerinden daha yüksek bir değerin sonuçları iyileştireceği, yanlış pozitif oranını düşüreceği gibi varsayımlarla belirlenmesidir. Genel olarak Psikometrik Tabanlı Yaklaşımlar için mevcut istatistiklerde kullanılan kesme noktaları, yeniden örnekleme sayısı, önyükleme sayısı ve anlamlılık düzeyi gibi değişkenler önceki çalışmalara göre daha optimum sonuçlar ortaya koyulmasını sağlamıştır. Bunlara alternatif olarak mevcut göstergelerin birlikte ele alınmasının oldukça etkili olduğu bu çalışmada saptanmıştır. Özellikle kullanılan gösterge sayısı ve yöntem sınıfı çeşitliliği arttıkça test hilesi gerçekleştiren bireylerin daha yüksek doğrulukta tespit edildiği gözlemlenmiştir. Daha önce bu bakış açısına benzer biçimde gerçekleştirilen araştırmalarda özellikle özgüllük oranları mevcut çalışmadaki sonuçların çok altında kalmıştır (Meijer, 2003; Steinkamp, 2017; Meijer ve Sijtsma, 2001). Önerilen bu yöntemde istatistiklerin kuramsal temelleri ve özellikle tespit ettikleri hile yapan birey kümeleri farklılaştığı için yüksek tespit oranları sağlanmıştır. Bu bakış açısı sadeliği yanında oldukça işlevsel bir bakış açısıdır.

Klasik yaklaşımlar ikili madde yanıtları, kategorik madde yanıtları ve tepki süreleri gibi girdi değişkenleri ile çalışmaktadır. Çalışmada özellikle sınav sürecinde elde edilen tüm bilgileri değerlendirmeye alarak ve çeşitli girdi değişkenleri türeterek sınav verilerinden en iyi şekilde yararlanmaya odaklanılmıştır. Bu bağlamda önerilen ilk iki yöntem birbirinden farklı yöntem sınıflarında bulunan ve farklı girdiler ile test hilesini tespit etme yaklaşımlarını tüm değişkenleri irdeleyerek ve ortak bir biçimde değerlendirmeye dahil ederek ele almıştır. Bu yaklaşımlar daha önce gerçekleştirilen araştırmalardan daha iyi tespit oranları sağlamıştır (Cizek ve Wollack, 2017; Ferrara, 2017; Kingston ve Clark, 2014; Man ve diğ., 2019; Sinharay ve diğ., 2017; van der Linden ve Guo, 2008; van der Linden ve Lewis, 2015; van der Linden ve Sotaridona, 2006; van der Linden, 2009; Wollack ve diğ., 2015; Wollack ve Fremer, 2013). Yöntemlerin birleşik kullanımında, göz önüne alınan farklı türde girdi değişkenlerinin kullanılması ve farklı istatistiklerin ortak gücü, test hilesinin belirlenmesi noktasında ümit vaad edicidir.

Bir diğer husus, sınav yöneticileri ve sınav sonrası hile tespiti ile ilgili soruşturma gerçekleştirecek araştırmacılar için mevcut çalışmanın odağında olan çift yönlü bakış açısıdır. İlki bir test senaryosunda, hileye ilişkin bir kaygı olması halinde çok yüksek sayılarda sınav

katılımcısının gözden geçirilmesinin mümkün olmadığı veya maliyetinin oldukça yüksek olduğu durumlarda şaibeli sınav için çalışmada önerilen yöntemlerin bir kısmı ile incelenecek katılımcı alt grubunun belirlenmesi; ikincisi hakkında gözlemsel yöntemler ile hileli davrandığı yönünde yargı olan bireyler hakkında etiketlemeyi sağlamlaştırmak ve gerekli önlemleri almak için nicel ek kanıt elde etme yönleri ile uygulayıcılar için daha işlevsel bir yönlendirme sağlayabilir.

Gerçekleştirilen araştımada Psikometrik Tabanlı Yaklaşımların tespit güçlerinin benzer çalışmalarla yakınsadığı görülmektedir. Çalışmada önerilen Yöntemlerin Birleşik Kullanımı Yaklaşımı ve Sınav Süreci Yaklaşımlarının duyarlılık, özgüllük ve genel doğruluk oranlarını oldukça artırdığı görülmektedir. Bunun yanında Denetimli Sınıflandırma Yöntemlerinin test hilesinin belirlenmesi noktasında işlevselliği sınanmış daha fazla girdi değişkenin etkili bir özellik seçimi ile tespit gücünü iyileştirebileceği belirlenmiştir. Özellikle daha evvel ilgili çalışmalarda kullanılmayan türetilmiş değişkenlerin girdi değişkenleri olarak çalışma kapsamına alınması önceki çalışmalara göre test hilesi gerçekleştiren bireylerin daha yüksek bir doğruluk oranı ile belirlenmesini sağlamıştır (Man ve diğ., 2019; Zopluoglu, 2019; Wang ve diğ., 2017). Ayrıca, denetimli yöntemler ile aynı sade mantık çalıştırıldığında tespit oranlarının daha iyi bir seviyeye ulaştığı görülmektedir. Alanda test hilesi ile ilgili çalışmaların büyük bir çoğunluğunun yeni istatistikler üretmek ve klasik yaklaşımlara odaklanmak olduğu gerçeği göz önüne alındığında denetimli yöntemlerin ve klasik yöntemlerin birleşik gücünden daha sade yollar ile yararlanmanın verimliliği ve etkililiği bu çalışma ile okuyuculara sunulmuştur.

Çalışmada kullanılan yaklaşımlar testlerde hileli davranışları birey düzeyinde ele almaktadır. Ancak bu çalışma kapsamında önerilen sınav süreci modelinin bir getirisi de, özellikle incelenecek sınav merkezlerinin belirlenmesi açısından grup veya merkezler düzeyinde de hile tespiti için önerilen yöntemlere alternatif sağlayabileceği görüşüdür. Bu açıdan hesaplama maliyetinden tasarruf ederek ayrıca uygulamaya ve araştırmacılara kolay ve işlevsel bir görüş sağlayabilir.

Eğitsel ölçüm ve özellikle çalışmanın konu alanı olan testlerde hile davranışlarının incelenmesinde makine öğrenmesi yöntemlerinin kullanımı oldukça az sayıda araştırmayla sınırlıdır (Man ve diğ., 2019; Wang ve diğ., 2017; Zopluoglu, 2019). Bunun yanında eğitimde makine öğrenmesi algoritmalarının kullanımı giderek daha yaygın hale gelmektedir. Çalışmanın odak noktası olan değerlendirme sonrası adli bilişim yoluyla test hilesinin tespiti

noktasında makine öğrenmesi yöntemlerinin kullanılmasının çeşitlendirilmesi ve farklı ölçüler ile türetilmiş değişkenlerin işe katılması açısından alan çalışmalarına çeşitlilik kazandıracağı düşünülmektedir.

Yine benzer şekilde test hilesi araştırmalarında makine öğrenmesi yöntemleri nadiren kullanılsa da makine öğrenmesi algoritmalarının birleşik ve bütünsel gücünden yararlanma üzerine henüz gerçekleştirilmiş bir çalışma bulunmamaktadır. Bu yöntemler temel olarak; algoritma artırma (boosting algorithms), torbalama algoritmaları (bagging algorithms) ve istifleme algoritmalarıdır (stacking algorithms). Önerilen basit topluluk oyu ve birleşik kullanım modelleri oldukça basit bir bakış açısıyla hile belirlemede tespit gücünün artırılması konusunda iyi sonuçlar vermiştir ve diğer yöntemlerde olduğu gibi uygulama kolaylığı açısından gayet işlevseldir.

5.2. Öneriler

Bu araştırma sonucunda ortaya çıkan sonuçlar ışığında araştırma ve uygulama boyutunda çeşitli öneriler okuyuculara sunulabilir:

- Bu tez çalışmasında Cizek ve Wollack'ın (2017) tarafından sunulan gerçek ve hazır bir veri seti kullanılmıştır. Benzer veri setleri ile araştırmada kullanılan etkili ve işlevsel yöntemler başarılı sonuçlar ortaya koyabilir.

- Araştırmada önerilen ilk yöntem olan göstergelerin birlikte kullanılması boyutunda Kişi-Uyum İstatistiklerinden l_z^*, CUSUM ve ZU3 Kişi-Uyum İstatistikleri; Kopya Ve Benzerlik Ölçülerinden K ve M4 İndeksleri; Tepki Süresi Modellerinden LNMRT ve χpf İstatistikleri kullanılmıştır. Aynı bakış açısı her bir yöntem gurubu için farklı indeks ve istatistikleri içerecek biçimde çalışılabilir.

- Mevcut çalışmada veri setindeki hazır değişkenlerden elde edilen çok sayıda türetilmiş değişken (kişi hızları, genel başarı, kişi yetenekleri gibi) kullanılmıştır. Sonraki araştırmalarda bu değişkenler çeşitlendirebilir ve artırılabilir.

- Denetimli Sınıflandırma Yöntemleri için KNN Sınıflandırıcısı, Naive Bayes Sınıflandırıcısı ve Rastgele Ormanlar Sınıflandırıcısı kullanılmıştır. Bu algoritmaların dışında farklı makine öğrenmesi algoritmaları sonraki araştırmacılar tarafından değerlendirmeye

alınabilir.

- Test hilesini tespit etmede makine öğrenmesi algoritmalarının birleşik ve bütünsel gücünden yararlanma noktasında önerilen yaklaşımlar haricinde daha kapsamlı ve çok boyutlu algoritma artırma (boosting algorithms), torbalama algoritmaları (bagging algorithms) ve istifleme algoritmalarıdır (stacking algorithms) gibi yöntemler sonraki araştırmalarda çalışılabilir.

- Uygulama kolaylığı açısından mevcut araştırma prosedürü ve önerilen yöntemleri kapsayan bir paket program uygulayıcılara sunulabilir.

KAYNAKÇA

Al-Fahoum, A. S., & Al-Fraihat, A. A. (2014). Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. *International Scholarly Research Notices, 2014*.

American Educational Research Association. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Anderson, J. A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society: Series B (Methodological), 46*(1), 1-22.

Archer, K. J., & Kimes, R. V. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, *52*(4), 2249-2260.

Armstrong, R. D., & Shi, M. (2009). Model-free CUSUM methods for person fit. *Journal of Educational Measurement*, *46*(4), 408-428.

Armstrong, R. D., Stoumbos, Z. G., & Shi, M. (2007). On the performance of the l z person-fit statistic. *Practical Assessment, Research, and Evaluation*, *12*(1), 16.

Ayala, R. J. (2009). The Theory and Practice of Item Response Theory. *New York: Guilford Publications Incorporated*.

Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical transactions of the Royal Society of London*, 1(53), 370-418.

Belov, D. I. (2013). Detection of test collusion via Kullback–Leibler divergence. *Journal of Educational Measurement*, *50*(2), 141-163.

Berkhin, P. (2006). A survey of clustering data mining techniques. *Grouping multidimensional data: Recent advances in clustering*, 25-71.

Bolsinova, M., & Tijmstra, J. (2018). Improving precision of ability estimation: Getting more

from response times. *British Journal of Mathematical and Statistical Psychology*, *71*(1), 13-38.

Boughton, K. A., Smith, J., & Ren, H. (2017). Using response time data to detect compromised items and/or people. In *Handbook of quantitative methods for detecting cheating on tests* (pp. 177-190). Routledge.

Bradlow, E. T., Weiss, R. E., & Cho, M. (1998). Bayesian identification of outliers in computerized adaptive tests. *Journal of the American Statistical Association*, *93*(443), 910-919.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123-140.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5-32.

Ceci, M. (2005). Naive bayesian learning from structural data. *Unpublished Doctoral Thesis), Computer Science and Engineering, University of Bari, Italy*.

Challagulla, V. U. B., Bastani, F. B., Yen, I. L., & Paul, R. A. (2005). Empirical assessment of machine learning based software defect prediction techniques. *International Journal on Artificial Intelligence Tools*, *17*(02), 389-400.

Chomboon, K., Chujai, P., Teerarassamee, P., Kerdprasop, K., & Kerdprasop, N. (2015, March). An empirical study of distance metrics for k-nearest neighbor algorithm. In *Proceedings of the 3rd international conference on industrial application engineering* (Vol. 2).

Cizek, G. J. (2012). Defining and distinguishing validity: interpretations of score meaning and justifications of test use. *Psychological methods*, *17*(1), 31.

Cizek, G. J., & Wollack, J. A. (Eds.). (2017). *Handbook of quantitative methods for detecting cheating on tests*. Abingdon, England: Routledge.

Clark, S. J., & Desharnais, R. A. (1998). Honest answers to embarrassing questions: Detecting cheating in the randomized response model. *Psychological methods*, *3*(2), 160.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on*

information theory, *13*(1), 21-27.

Cutler, A., Cutler, D. R., & Stevens, J. R. (2011). Random forests. Ensemble machine learning. In *Ensemble Machine Learning* (pp. 157-175).

De La Torre, J., & Deng, W. (2008). Improving person-fit assessment by correcting the ability estimate and its reference distribution. *Journal of Educational Measurement*, *45*(2), 159-177.

Drasgow, F., Levine, M. V., & McLaughlin, M. E. (1991). Appropriateness measurement for some multidimensional test batteries. *Applied Psychological Measurement*, *15*(2), 171-191.

Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, *38*(1), 67-86.

Drasgow, F., Levine, M. V., & Zickar, M. J. (1996). Optimal identification of mismeasured individuals. Applied Measurement in Education, 9, 47–64.

Emons, W. H., Meijer, R. R., & Sijtsma, K. (2002). Comparing simulated and theoretical sampling distributions of the U3 person-fit statistic. *Applied Psychological Measurement*, *26*(1), 88-108.

Ferrara, S. (2017). A framework for policies and practices to improve test security programs: Prevention, detection, investigation, and resolution (PDIR). *Educational Measurement: Issues and Practice*, *36*(3), 5-23.

Fix, E., & Hodges, J. L. (1951). Nonparametric discrimination: consistency properties. Randolph Field, Texas, Project, 21-49.

Fox, J. P., & Marianti, S. (2017). Person-fit statistics for joint models for accuracy and speed. *Journal of Educational Measurement*, *54*(2), 243-262.

Fox, J. P., Klotzke, K., & Simsek, A. S. (2021). LNIRT: An R package for joint modeling of response accuracy and times. *arXiv preprint arXiv:2106.10144*.

Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. Elsevier.

Glas, C. A. W., & Dagohoy, A. V. T. (2007). A person fit test for IRT models for polytomous items. *Psychometrika, 72*, 159-180.

Glas, C. A., & van der Linden, W. J. (2010). Marginal likelihood inference for a model for item responses and response times. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 603-626.

Gordan, M., Georgakis, A., Tsatos, O., Oltean, G., & Miclea, L. (2006, May). Computational complexity reduction of the support vector machine classifiers for image analysis tasks through the use of the discrete cosine transform. In *2006 IEEE International Conference on Automation, Quality and Testing, Robotics* (Vol. 2, pp. 350-355). IEEE.

Granitto, P. M., Furlanello, C., Biasioli, F., & Gasperi, F. (2006). Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometrics and intelligent laboratory systems*, *83*(2), 83-90.

Gupta, P. (2017). Decision trees in machine learning. *Towards Data Science*, *17*.

Guttman, L. (1944). A basis for scaling qualitative data. *American sociological review*, *9*(2), 139-150.

Hanson, B. A., Harris, D. J., & Brennan, R. L. (1987). *A comparison of several statistical methods for examining allegations of copying*. Iowa City, IA: American College Testing Program.

Harnisch, D. L. (1983). Item response patterns: Applications for educational practice. *Journal of Educational Measurement*, 191-206.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (p. 359). New York: Springer-Verlag New York.

Hauser, C., Kingsbury, G. G., & Houser, R. L. (2011, April). Individual Score Validity: Using the Wariness Index to Identify Test Performance to Treat with Caution. In *Presented at the Annual Meeting of the National Council on Measurement in Education*.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, *20*(8), 832-844.

Holland, P. W. (1996). Assessing unusual agreement between the incorrect answers of two examinees using the K-index: Statistical theory and empirical support. *ETS Research Report Series*, *1996*(1), i-41.

Hong, M., & Can, S. (2017). Package 'PsyControl': CUSUM Person Fit Statistics.

Huang, T. W. (2012). Aberrance detection powers of the BW and Person-Fit Indices. *Journal of Educational Technology & Society*, *15*(1), 28-37.

James, J., Sandhya, L., & Thomas, C. (2013, December). Detection of phishing URLs using machine learning techniques. In *2013 International conference on control communication and computing (ICCC)* (pp. 304-309). IEEE.

Jatana, N., & Sharma, K. (2014, March). Bayesian spam classification: Time efficient radix encoded fragmented database approach. In *2014 International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 939-942). IEEE.

Juszczak, P., Tax, D., & Duin, R. P. (2002, May). Feature scaling in support vector data description. In *Proc. asci* (pp. 95-102). Citeseer.

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, *16*(4), 277-298.

Kingston, N., & Clark, A. (Eds.). (2014). *Test fraud: Statistical detection and methodology*. Routledge.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., ... & Team, R. C. (2020). Package 'caret'. *The R Journal*, *223*(7).

Lantz, B. (2013). Probabilistic Learning Classification Using Naive Bayes. Machine Learning with R, Jones, J.(ed.), 99-110.

Levine, M. V., & Drasgow, F. (1988). Optimal appropriateness measurement. *Psychometrika*, *53*(2), 161-176.

Levine, M. V., & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of educational statistics, 4*(4), 269-290.

Li, M. N. F., & Olejnik, S. (1997). The power of Rasch person-fit statistics in detecting unusual response patterns. *Applied Psychological Measurement, 21*(3), 215-231.

Man, K., Harring, J. R., & Sinharay, S. (2019). Use of data mining methods to detect test fraud. *Journal of Educational Measurement, 56*(2), 251-279.

Marianti, S., Fox, J. P., Avetisyan, M., Veldkamp, B. P., & Tijmstra, J. (2014). Testing for aberrant behavior in response time modeling. *Journal of Educational and Behavioral Statistics, 39*(6), 426-451.

Maynes, D. (2014). Detection of non-independent test taking by similarity analysis. In *Test Fraud* (pp. 69-96). Routledge.

Maynes, D. D. (2017). Detecting potential collusion among individual examinees using similarity analysis. In *Handbook of quantitative methods for detecting cheating on tests* (pp. 47-69). Routledge.

McLeod, L., Lewis, C., & Thissen, D. (2003). A Bayesian method for the detection of item preknowledge in computerized adaptive testing. *Applied Psychological Measurement, 27*(2), 121-137.

Meijer, R. R. (1996). Person-fit research: An introduction. *Applied Measurement in Education, 9*(1), 3-8.

Meijer, R. R. (1997). Person fit and criterion-related validity: An extension of the Schmitt, Cortina, and Whitney study. *Applied Psychological Measurement, 21*(2), 99-113.

Meijer, R. R. (2003). Diagnosing item score patterns on a test using item response theory-based person-fit statistics. *Psychological Methods, 8*(1), 72.

Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied psychological measurement, 25*(2), 107-135.

Meijer, R. R., & Sotaridona, L. (2006). Detection of advance item knowledge using response

times in computer adaptive testing.

Meijer, R. R., & Tendeiro, J. N. (2012). The use of the lz and lz* person-fit statistics and problems derived from model misspecification. *Journal of Educational and Behavioral Statistics*, *37*(6), 758-766.

Menzies, T., Greenwald, J., & Frank, A. (2006). Data mining static code attributes to learn defect predictors. *IEEE transactions on software engineering*, *33*(1), 2-13.

Menzies, T., Turhan, B., Bener, A., Gay, G., Cukic, B., & Jiang, Y. (2008, May). Implications of ceiling effects in defect predictors. In *Proceedings of the 4th international workshop on Predictor models in software engineering* (pp. 47-54).

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. C., ... & Meyer, M. D. (2019). Package 'e1071'. *The R Journal*.

Molenaar, I. W., & Hoijtink, H. (1990). The many null distributions of person fit indices. *Psychometrika*, *55*, 75-106.

Molenaar, I. W., & Hoijtink, H. (1996). Person-fit and the Rasch model, with an application to knowledge of logical quantors. *Applied Measurement in Education*, *9*(1), 27-45.

Montgomery, D. C. (2012). *Statistical quality control* (Vol. 7). New York: Wiley.

Mousavi, A., Cui, Y., & Rogers, T. (2019). An examination of different methods of setting cutoff values in person fit research. *International Journal of Testing*, *19*(1), 1-22.

Munson, J. C., & Khoshgoftaar, T. M. (1990). Regression modelling of software quality: empirical investigation. *Information and Software Technology*, *32*(2), 106-114.

Nering, M. L. (1995). The distribution of person fit using true and estimated person parameters. *Applied Psychological Measurement*, *19*(2), 121-129.

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Detecting careless respondents in web-based questionnaires: Which method to use?. *Journal of Research in Personality*, *63*, 1-11.

Padberg, F., Ragg, T., & Schoknecht, R. (2004). Using machine learning for estimating the defect content after an inspection. *IEEE Transactions on Software Engineering*, *30*(1), 17-28.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International journal of remote sensing*, *26*(1), 217-222.

Panda, M., & Patra, M. R. (2007). Network intrusion detection using naive bayes. *International journal of computer science and network security*, *7*(12), 258-263.

Partchev, I., Partchev, M. I., & Suggests, M. A. S. S. (2017). Package 'irtoys'. *A collection of functions related to item response theory (IRT)*.

Patton, J. M., Cheng, Y., Hong, M., & Diao, Q. (2019). Detection and treatment of careless responses to improve item parameter estimation. *Journal of Educational and Behavioral Statistics*, *44*(3), 309-341.

Qian, H., Staniewska, D., Reckase, M., & Woo, A. (2016). Using response time to detect item preknowledge in computer-based licensure examinations. *Educational Measurement: Issues and Practice*, *35*(1), 38-47.

RColorBrewer, S., & Liaw, M. A. (2018). Package 'randomforest'. *University of California, Berkeley: Berkeley, CA, USA*.

Reise, S. P. (1990). A comparison of item-and person-fit methods of assessing model-data fit in IRT. *Applied Psychological Measurement*, *14*(2), 127-137.

Reise, S. P. (1995). Scoring method and the detection of person misfit in a personality assessment context. *Applied Psychological Measurement*, *19*(3), 213-229.

Reise, S. P., & Widaman, K. F. (1999). Assessing the fit of measurement models at the individual level: A comparison of item response theory and covariance structure approaches. *Psychological Methods*, *4*(1), 3.

Ripley, B., Venables, W., & Ripley, M. B. (2015). Package 'class'. *The Comprehensive R Archive Network*, *11*.

Rizopoulos, D., & Rizopoulos, M. D. (2018). Package 'ltm'. *URL http://wiki. r-project. org/rwiki/doku. php*.

Roman, V. (2018). Machine Learning Introduction: *A Comprehensive Guide*.

Rupp, A. A. (2013). A systematic review of the methodology for person fit research in item response theory: Lessons about generalizability of inferences from the design of simulation studies. *Psychological Test and Assessment Modeling*, *55*(1), 3.

Sarkar, D., Sarkar, M. D., & KernSmooth, S. (2015). Package 'lattice'. *Version 0.20, 33*.

Schnikpe, D. L., & Scrams, D. J. (1999). Exploring Issues of Test Taker Behavior: Insights Gained from Response-Time Analyses. Law School Admission Council Computerized Testing Report. LSAC Research Report Series.

Seo, D. G., & Weiss, D. J. (2013). lz Person-fit index to identify misfit students with achievement test data. *Educational and Psychological Measurement*, *73*(6), 994-1016.

Sijtsma, K., & Meijer, R. R. (1992). A method for investigating the intersection of item response functions in Mokken's nonparametric IRT model. *Applied Psychological Measurement*, *16*(2), 149-157.

Sinharay, S. (2016). Asymptotically correct standardization of person-fit statistics beyond dichotomous items. *Psychometrika*, *81*(4), 992-1013.

Sinharay, S. (2017). Detection of item preknowledge using likelihood ratio test and score test. *Journal of Educational and Behavioral Statistics*, *42*(1), 46-68.

Sinharay, S. (2018). A new person-fit statistic for the lognormal model for response times. *Journal of Educational Measurement*, *55*(4), 457-476.

Sinharay, S. (2020). Detection of item preknowledge using response times. *Applied Psychological Measurement*, *44*(5), 376-392.

Sinharay, S. (2021). Latent-variable approaches utilizing both item scores and response times to detect test fraud. *Open Education Studies*, *3*(1), 1-16.

Sinharay, S., & Johnson, M. S. (2020). The use of item scores and response times to detect examinees who may have benefited from item preknowledge. *British Journal of Mathematical and Statistical Psychology*, *73*(3), 397-419.

Sinharay, S., Duong, M. Q., & Wood, S. W. (2017). A new statistic for detection of aberrant answer changes. *Journal of Educational Measurement*, *54*(2), 200-217.

Snijders, T. A. (2001). Asymptotic null distribution of person fit statistics with estimated person parameter. *Psychometrika*, *66*, 331-342.

Sotaridona, L. S., & Meijer, R. R. (2002). Statistical properties of the K-index for detecting answer copying. *Journal of Educational Measurement*, *39*(2), 115-132.

Sotaridona, L. S., & Meijer, R. R. (2003). Two new statistics to detect answer copying. *Journal of Educational Measurement*, *40*(1), 53-69.

Steinkamp, S. C. (2017). *Identifying Aberrant Responding: Use of Multiple Measures* (Doctoral dissertation, University of Minnesota).

Stigler, S. M. (1982). Thomas Bayes's bayesian inference. *Journal of the Royal Statistical Society: Series A (General)*, *145*(2), 250-258.

St-Onge, C., Valois, P., Abdous, B., & Germain, S. (2011). Accuracy of person-fit statistics: A Monte Carlo study of the influence of aberrance rates. *Applied Psychological Measurement*, *35*(6), 419-432.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323.

Tatsuoka, K. K. (1984). Caution indices based on item response theory. *Psychometrika*, *49*, 95-110.

Tatsuoka, K. K., & Tatsuoka, M. M. (1983). Spotting erroneous rules of operation by the individual consistency index. *Journal of Educational Measurement*, *20*(3), 221-230.

Tax, D. M., & Duin, R. P. (2002, August). Using two-class classifiers for multiclass

classification. In *2002 International Conference on Pattern Recognition* (Vol. 2, pp. 124-127). IEEE.

Team, R. D. C. (2023). A language and environment for statistical computing. *http://www. R-project. org*.

Tendeiro, J. N., Meijer, R. R., & Niessen, A. S. M. (2016). PerFit: An R package for person-fit analysis in IRT. *Journal of Statistical Software*, *74*, 1-27.

Thiessen, B. A. (2008). *Relationship between test security policies and test score manipulations* (Doctoral dissertation, University of Iowa).

Turhan, B., & Bener, A. B. (2007, July). Software Defect Prediction: Heuristics for Weighted Naïve Bayes. In *ICSOFT (SE)* (pp. 244-249).

Van der Flier, H. (1980). *Vergelijkbaarheid van individuele testprestaties*. Swets & Zeitlinger.

Van Der Flier, H. (1982). Deviant response patterns and comparability of test scores. *Journal of Cross-Cultural Psychology*, *13*(3), 267-298.

Van Der Flier, H., & Drenth, P. J. (1977). Fair Selection and Comparability of Testscores.

Van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*(3), 287.

Van Der Linden, W. J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement*, *46*(3), 247-272.

Van der Linden, W. J., & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika*, *73*(3), 365-384.

Van der Linden, W. J., & Lewis, C. (2015). Bayesian checks on cheating on tests. *Psychometrika*, *80*(3), 689-706.

Van der Linden, W. J., & Sotaridona, L. (2006). Detecting answer copying when the regular response process follows a known response model. *Journal of Educational and Behavioral Statistics*, *31*(3), 283-304.

Van Krimpen-Stoop, E. M., & Meijer, R. R. (1999). The null distribution of person-fit statistics for conventional and adaptive tests. *Applied Psychological Measurement*, *23*(4), 327-345.

Van Krimpen-Stoop, E. M., & Meijer, R. R. (2001). CUSUM-based person-fit statistics for adaptive testing. *Journal of Educational and Behavioral Statistics*, *26*(2), 199-217.

Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology*, *68*(3), 456-477.

Wang, X., Liu, Y., & Hambleton, R. K. (2017). Detecting item preknowledge using a predictive checking method. *Applied psychological measurement*, *41*(4), 243-263.

Watts, J. D., Powell, S. L., Lawrence, R. L., & Hilker, T. (2011). Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery. *Remote Sensing of Environment*, *115*(1), 66-75.

Wise, S. L., Ma, L., Kingsbury, G. G., & Hauser, C. (2010). An Investigation of the Relationship between Time of Testing and Test-Taking Effort. *Northwest Evaluation Association*.

Wollack, J. A. (1996). *Detection of answer copying using item response theory*. The University of Wisconsin-Madison.

Wollack, J. A. (2003). Comparison of answer copying indices with real data. *Journal of Educational Measurement*, *40*(3), 189-205.

Wollack, J. A. (2006). Simultaneous use of multiple answer copying indexes to improve detection rates. *Applied Measurement in Education*, *19*(4), 265-288.

Wollack, J. A., & Cohen, A. S. (1998). Detection of answer copying with unknown item and trait parameters. *Applied Psychological Measurement*, *22*(2), 144-152.

Wollack, J. A., & Eckerly, C. A. (2016). Detecting test tampering at the group level. In *Handbook of quantitative methods for detecting cheating on tests* (pp. 214-231). Routledge.

Wollack, J. A., & Fremer, J. J. (Eds.). (2013). *Handbook of test security*. Routledge.

Wollack, J. A., & Schoenig, R. W. (2018). *Cheating*. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 260–265). Thousand Oaks, CA: Sage.

Wollack, J. A., Cohen, A. S., & Eckerly, C. A. (2015). Detecting test tampering using item response theory. *Educational and Psychological Measurement, 75*(6), 931-953.

Zopluoglu, C. (2019). Computation of the Response Similarity Index M4 in R under the Dichotomous and Nominal Item Response Models. *International Journal of Assessment Tools in Education, 6*(5), 1-19.

Zopluoglu, C., & Zopluoglu, M. C. (2018). Package 'CopyDetect'.

EKLER

Ek-1: Çalışmada Kullanılan R Kodları

#Kişi-Uyum İstatistikleri İçin R Kodları

```
> library(PerFit)

> set.seed(123)

> yanıtlarikili<-read.table("yanıtlarikili.txt",header=T)

> lzstarout<-lzstar(yanıtlarikili,NA.method = "Pairwise", Save.MatImp = FALSE,

      IP = NULL, IRT.PModel = "2PL", Ability = NULL, Ability.PModel = "ML")

> lzstarcut<-cutoff(lzstarout,Blvl = 0.05, Breps = 10000)

> lzstarflag<-flagged.resp(lzstarout,cutoff.obj=lzstarcut,scores=F)

> ZU3out<-ZU3(yanıtlarikili,NA.method = "Pairwise", Save.MatImp = FALSE,

      IP = NULL, IRT.PModel = "2PL", Ability = NULL, Ability.PModel = "ML")

> ZU3cut<-cutoff(ZU3out,ModelFit = "NonParametric", Nreps=1000,

            IP=x$IP, IRT.PModel=x$IRT.PModel, Ability=x$Ability,

            Ability.PModel=x$Ability.PModel, mu=0, sigma=1,

            Blvl = 0.05, Breps = 10000)

> ZU3flag<-flagged.resp(ZU3out,cutoff.obj=ZU3cut,scores=F)

> library(PsyControl)

> CUSUMout<-cusum(dat=yanıtlarikili)

> CUSUMcut<-cusum.cutoff(CUSUMout,upp=.975,low=.025,Breps=1000)

> CUSUMflag<-cusum.flag(CUSUMout,CUSUMcut)
```

```
> lzstarcıktı<-lzstarflag$PFSscores[,1]

> ZU3cıktı<-ZU3flag$PFSscores[,1]

> kopyacılar<-
c(2,3,4,5,11,103,104,141,149,191,219,226,248,251,271,292,415,452,453,478,494,498,505,52
4,623,624,639,645,687,892,903,919,921,923,983,1176,1267,1307,1356,1358,1413,1426,1432
,1450,1585,1620)

> length(lzstarcıktı)

> length(U3cıktı)

> intersect(lzstarcıktı,kopyacılar)

> intersect(ZU3cıktı,kopyacılar)

> CUSUMflag

>CUSUMcıktı<-
c(1,5,9,12,14,17,28,39,49,77,80,107,108,122,126,129,154,175,200,213,245,251,328,337,339,
345,371,410,412,419,423,439,502,518,525,624,655,704,724,746,977,1027,1032,1084,1138,1
148,1171,1192,1193,1214,1265,1346,1496,1633)

> length(CUSUMcıktı)

> intersect(CUSUMcıktı,kopyacılar)
```

#Kopya Ve Benzerlik Analizleri İçin R Kodları

```
>yanıtlar<-read.table("yanıtlarnominal.txt",heder=TRUE)

>demografik<-read.table("demografik.txt",header=TRUE)

>demografik[,2:8]<-NULL

>demografik<-[,3]<-NULL

>data<-cbind(demografik,yanıtlar)
```

>kopyacılar<-c(e100002, e100003, e100004, e100005, e100011, e100103, e100104, e100141, e100149, e100191, e100219, e100226, e100248, e100251, e100271, e100292, e100415, e100452, e100453, e100478, e100494, e100498, e100505, e100524, e100623, e100624, e100639, e100645, e100687, e100892, e100903, e100919, e100921, e100923, e100983, e101176, e101267, e101307, e101356, e101358, e101413, e101426, e101432, e101450, e101585, e101620)

R console ile veri aktarmak için;

>merkezler<-subset(CredentialForm1,select="Cent_id")

>merkezler<-union(merkezler,merkezler)

>kop<-similarity1(data=data, model="1PL", person.id="EID", center.id="Cent_id", item.loc=3:172, centers=c("#sırasıyla merkezler"))

#Çıktı kümeleri;

>M405<-
c(740,222,168,1166,1595,57,1601,653,1193,288,848,248,505,624,199,198,225,304,452,431,523,191,421,380,381,389,157,1196,498,1601,653,1311,98,201,210,455,85,1134,144,214,277,941,605,1246,846,415,364,607,233,721,1469,558,408,398,917,762,685,651,750,1123,531,218,379,319,535,1049,446,497,266,568,149,213,1295,1336,1114,1474,510,672,474,127,107,1160,109,990,451,857,910,524,909,667,1587,370,1002,884,1536,1419,820,965,1090,172,747,1618,1449,285,189,1627,839,255,205,862,752,1572,803,1391,1442,694,608,1484,1121,1597,1624,776,334,1329,237,797,238,1460,860,1007,891,904,1468,444,1563,1527,901,738,343,1080,890,626,643,226,577,372,428,1502,1181,927,1519,1376,851,461,802,835,623,921,766,1345,821,1401,1549,1559,539,587,1173,1025,488,1450,1347,1546,1516,487,339,1544,616,1183,634,1014,640,1425,1302,1574,916,1290,1217,1566,824,669,512,1576,1200,666,600,926,953,1596,588,986,686,652,822,612,359,1501,786,1571,335,330,893,897,676,1213,843,1222,1102,993,1584,833,1498,1379,1459,1577,1163,1558,1199,1374,917,1348,316,1348,1417,1326,1455,944,1339,1358,281,463,287,1477,560,1521,1088,861,540,1162,838,551,525,875,434,691,961,406,1383,436,827,438,1415,699,976,596,1342,1359,1004,1540,674,660,1510,1626,1475,1464,1413,870,1170,486,696,1318,1176,549,793,1369,1285,701,1245,593,1185,773,1446,923,899,720,1398,1321,980,968,789,778,1538,963,1452,1062,1356,1024,942,1491,1430,1276,10

79,1555,1560,1629,1505,1503,1620,1437,1582,1454,1415,699,976,596,1342,1359,1004,1540
,674,660,1510,1626,1475,1464,1413,870,1170,486,696,1318,1607,491,1305,1066,1201,1011,
902,350,166,338,1093,1567,464,585,122,81,135,193,1403,1141,453,1084,249,619,274,352,4
66,245,494,399,974,889,842,566,481,93,1207,1092,779,787,1335,1631,128,420,1499,1300,7
00)

>K05<-
c(222,248,505,624,199,225,304,452,431,523,191,98,85,201,210,455,607,233,408,398,762,11
23,218,685,379,535,1049,446,568,213,1336,1114,474,990,909,1587,370,884,172,189,255,20
5,752,694,608,1484,1121,1597,1624,1007,891,1468,1080,890,226,577,1502,1376,587,487,15
44,640,916,1290,1566,824,669,926,588,986,953,359,786,1571,1501,893,833,1577,1379,1374
,316,1417,1326,281,287,560,1088,1162,838,551,525,875,691,1359,1170,1176,793,1285,1321
,1452,1024,942,1555,1276,1437,1454,1359,870,1170,1066,902,350,166,338,1093,464,122,45
3,619,453,494,399,481,128)

>kopyacılar<-
c(2,3,4,5,11,103,104,141,149,191,219,226,248,251,271,292,415,452,453,478,494,498,505,52
4,623,624,639,645,687,892,903,919,921,923,983,1176,1267,1307,1356,1358,1413,1426,1432
,1450,1585,1620)

> length(M405)

> intersect(kopyacılar,M405)

> length(intersect(kopyacılar,M405))

> length(K05)

> intersect(kopyacılar,K05)

> length(intersect(kopyacılar,K05))

#Tepki Süresi Modelleri İçin R Kodları

##LNMRT İçin R Kodları;

library(LNIRT)

```
data(CredentialForm1)
Y <- as.matrix(CredentialForm1[c(which(colnames(CredentialForm1)=="iraw.1")
   :which(colnames(CredentialForm1)=="iraw.170"))])
N <- nrow(Y)
RT<-as.matrix(CredentialForm1[c(which(colnames(CredentialForm1)=="idur.1")
    :which(colnames(CredentialForm1)=="idur.170"))])
RT[RT==0]<-NA
RT<-log(RT)
XFT <- data.frame(CredentialForm1[1:10],stringsAsFactors=TRUE) #Arka Plan Değişkenleri
XFT$Tot_time <- (XFT$Tot_time-mean(XFT$Tot_time))/sqrt(var(XFT$Tot_time))
XFT$Pgroup <- matrix(0,ncol=2,nrow=N)
XFT$Pgroup[XFT$Pretest==6,1] <- -1
XFT$Pgroup[XFT$Pretest==6,2] <- -1
XFT$Pgroup[XFT$Pretest==7,1] <- 1
XFT$Pgroup[XFT$Pretest==8,2] <- 1
XFT$Cgroup <- matrix(0,ncol=3,nrow=N)
XFT$Cgroup[XFT$Country=="USA",1] <- 1
XFT$Cgroup[XFT$Country=="Philippines",2] <- 1
XFT$Cgroup[XFT$Country=="India",3] <- 1
XFT$Cgroup[c(XFT$Country!="USA" & XFT$Country!="India" &
XFT$Country!="Philippines"),1:3] <- -1
```

```
 XA <- matrix(unlist(XFT[,c("Pgroup","Tot_time")]),ncol=3,nrow=N)

 XT <- matrix(unlist(XFT[,c("Pgroup")]),ncol=2,nrow=N)

 set.seed(123)

 out <- LNIRT(RT=RT,Y=Y,XG=5000,XGresid=10000,XPA=XA,XPT=XT,residual=TRUE)
```

##Çıktı listeleri;

```
Liste1<-
c(8,11,20,21,22,25,32,33,36,37,51,57,61,67,69,71,87,93,94,112,137,141,185,192,198,199,219
,226,240,248,249,258,265,290,294,308,311,323,337,350,356,400,410,414,439,448,455,466,4
78,480,494,498,505,524,542,550,567,580,587,591,612,624,638,645,661,670,678,679,682,702
,720,727,745,761,783,816,834,853,890,918,944,964,1001,1003,1011,1053,1057,1123,1147,1
153,1166,1198,1253,1279,1294,1295,1296,1297,1298,1304,1310,1317,1324,1370,1391,1406,
1410,1414,1434,1458,1465,1482,1507,1513,1536,1567,1591)

Liste2<-
c(543,598,631,666,677,679,681,694,695,698,699,701,710,711,713,714,716,719,826,872,933,
936,939,957,958,961,976,983,989,991,995,1006,1008,1017,1022,1307,1349,1364,1432,1479,
1489,1496,1502,1505,1521,1533,1539,1540,1541,1547,1569,1574,1576,1579,1590,1593,159
4,1599,1614,1620,1621,1622,1630,1631,1632,1634)

>kopyacılar<-
c(2,3,4,5,11,103,104,141,149,191,219,226,248,251,271,292,415,452,453,478,494,498,505,52
4,623,624,639,645,687,892,903,919,921,923,983,1176,1267,1307,1356,1358,1413,1426,1432
,1450,1585,1620)

birlesim<-union(Liste1,Liste2)

sonuc<-length(birlesim,kopyacılar)
```

χpf İçin R Kodları;

```
ChiRT=function(alpha,beta,RT)
```

```
{n=nrow(RT)

tauhat=rep(sum(alpha*alpha*beta)/sum(alpha*alpha),n)-
RT%*%(alpha*alpha)/(sum(alpha*alpha))

v=RT-rep(1,n)%*%t(beta)+tauhat%*%t(rep(1,length(beta)))

chisq=(v*v)%*%(alpha*alpha)

return(chisq)

}

> lt<-read.table("chisq.txt",header=T)

> lt<-edit(lt)

> lnrt<-subset(lt,a>=205)

##Çıktılar;

> lnrt<-list(lnrt)

> lnrt<-edit(lnrt)

> ln<-c("5",  "7", "8", "11", "20", "25", "28", "31", "36", "40", "52", "61",  "64", "69", "70",
"71", "85", "87", "91", "92", "106", "107",  "115", "122", "124", "126", "131", "137", "141",
"142", "145",  "149", "153", "185", "190", "198", "199", "219", "226", "237",  "240", "248",
"249", "251", "255", "258", "265", "266", "269",  "276", "281", "289", "291", "292", "300",
"302", "307", "308",  "311", "317", "324", "328", "344", "351", "356", "361", "375",  "377",
"383", "385", "395", "400", "402", "414", "423", "430",  "439", "448", "452", "455", "478",
"480", "483", "485", "494",  "498", "500", "520", "524", "525", "535", "542", "565", "567",
"570", "580", "581", "587", "591", "600", "612", "615", "624",  "626", "635", "638", "645",
"649", "678", "702", "722", "727",  "736", "745", "747", "750", "755", "756", "761", "768",
"772",  "783", "791", "793", "796", "803", "804", "823", "833", "834",  "847", "857", "884",
"890", "920", "924", "926", "930", "936",  "944", "950", "952", "964", "968", "987", "1001",
"1003", "1011", "1012", "1016", "1034", "1042", "1045", "1051", "1053", "1057", "1074",
```

"1089", "1090", "1104", "1115", "1120", "1123", "1145", "1147", "1150", "1153", "1158", "1164", "1175", "1188", "1195", "1200", "1210", "1213", "1214", "1225", "1241", "1242", "1247", "1252", "1253", "1264", "1294", "1295", "1297", "1310", "1324", "1335", "1339", "1348", "1367", "1375", "1391", "1399", "1406", "1407", "1412", "1414", "1420", "1434", "1438", "1446", "1447", "1449", "1459", "1465", "1466", "1480", "1482", "1484", "1497", "1501", "1504", "1513", "1533", "1536", "1537", "1540", "1549", "1563", "1579", "1581", "1598", "1618", "1622", "1626", "1631", "1634")

> kopyacılar<-
c(2,3,4,5,11,103,104,141,149,191,219,226,248,251,271,292,415,452,453,478,494,498,505,
524,623,624,639,645,687,892,903,919,921,923,983,1176,1267,1307,1356,1358,1413,1426,
1432,1450,1585,1620)

> length(intersect(kopyacılar,ln))

#Yöntemlerin Birleşik Kullanımı İçin R Kodları;

>read.table("lzstarcıktı.txt",header=T)

>read.table("birlesim.txt",header=T)

>read.table("ZU3cktı.txt",header=T)

>read.table("K05.txt",header=T)

>read.table("M405.txt",header=T)

>read.table("ln.txt",header=T)

>m1<-union(lzstarcıktı,birlesim)

>length(intersect(m1,kopyacılar)

>m2<-union(ZU3cıktı,K05)

>length(intersect(m2,kopyacılar)

>m3<-union(CUSUMcıktı,M405)

>length(intersect(m3,kopyacılar)

>m4<-union(ZU3cıktı,birlesim)

>length(intersect(m4,kopyacılar)

>m5<-union(birlesim,K05)

>length(intersect(m5,kopyacılar)

>m6<-union(M405,ln)

>length(intersect(m6,kopyacılar)

>m7<-union(m4,K05)

>length(intersect(m7,kopyacılar)

>B<-union(CUSUMcıktı,ln)

>m8<-union(B,K05)

>length(intersect(m8,kopyacılar)

>B1<-union(lzstarcıktı,K05)

>m9<-union(B1,ln)

>length(intersect(m9,kopyacılar)

>m10<-union(m2,ln)

>length(intersect(m10,kopyacılar)

>B2<-union(CUSUMcıktı,LNMRT)

>m11<-union(B2,M405)

>length(intersect(m11,kopyacılar)

#Sınav Süreci Modeli İçin R Kodları

##Devriyeler için kodlar;

###Kişi-Uyum Analizleri .01 kesme seviyesi için;

```
> library(PerFit)
> set.seed(123)
> yanıtlarikili<-read.table("yanıtlarikili.txt",header=T)
> lzstarout<-lzstar(yanıtlarikili,NA.method = "Pairwise", Save.MatImp = FALSE,
        IP = NULL, IRT.PModel = "2PL", Ability = NULL, Ability.PModel = "ML")
> lzstarcut<-cutoff(lzstarout,Blvl = 0.01, Breps = 10000)
> lzstarflag<-flagged.resp(lzstarout,cutoff.obj=lzstarcut,scores=F)
> ZU3out<-ZU3(yanıtlarikili,NA.method = "Pairwise", Save.MatImp = FALSE,
        IP = NULL, IRT.PModel = "2PL", Ability = NULL, Ability.PModel = "ML")
> ZU3cut<-cutoff(ZU3out,ModelFit = "NonParametric", Nreps=1000,
                IP=x$IP, IRT.PModel=x$IRT.PModel, Ability=x$Ability,
                Ability.PModel=x$Ability.PModel, mu=0, sigma=1,
                Blvl = 0.01, Breps = 10000)
> ZU3flag<-flagged.resp(ZU3out,cutoff.obj=ZU3cut,scores=F)
> library(PsyControl)
> CUSUMout<-cusum(dat=yanıtlarikili)
> CUSUMcut<-cusum.cutoff(CUSUMout,upp=.995,low=.005,Breps=1000)
> CUSUMflag<-cusum.flag(CUSUMout,CUSUMcut)
```

> lzstarcıktı<-lzstarflag$PFSscores[,1]

> ZU3cıktı<-ZU3flag$PFSscores[,1]

> kopyacılar<-
c(2,3,4,5,11,103,104,141,149,191,219,226,248,251,271,292,415,452,453,478,494,498,505,524,623,624,639,645,687,892,903,919,921,923,983,1176,1267,1307,1356,1358,1413,1426,1432,1450,1585,1620)

> length(lzstarcıktı)

> length(U3cıktı)

> intersect(lzstarcıktı,kopyacılar)

> intersect(ZU3cıktı,kopyacılar)

> CUSUMflag

>CUSUMcıktı<-c(1,14,17,49,126,251,339,410,1027,1496)

> length(CUSUMcıktı)

> intersect(CUSUMcıktı,kopyacılar)

##Bekçiler için kodlar;

###Kopya ve Benzerlik Analizleri .01 anlamlılık düzeyi için;

>yanıtlar<-read.table("yanıtlarnominal.txt",heder=TRUE)

>demografik<-read.table("demografik.txt",header=TRUE)

>demografik[,2:8]<-NULL

>demografik<-[,3]<-NULL

>data<-cbind(demografik,yanıtlar)

>kopyacılar<-c(e100002, e100003, e100004, e100005, e100011, e100103, e100104, e100141,

e100149, e100191, e100219, e100226, e100248, e100251, e100271, e100292, e100415, e100452, e100453, e100478, e100494, e100498, e100505, e100524, e100623, e100624, e100639, e100645, e100687, e100892, e100903, e100919, e100921, e100923, e100983, e101176, e101267, e101307, e101356, e101358, e101413, e101426, e101432, e101450, e101585, e101620)

R console ile veri aktarmak için;

>merkezler<-subset(CredentialForm1,select="Cent_id")

>merkezler<-union(merkezler,merkezler)

>kop<-similarity1(data=data, model="1PL", person.id="EID", center.id="Cent_id", item.loc=3:172, c>enters=c(“#sırasıyla merkezler”))

#Çıktı kümeleri;

M401<-
c(505,225,304,452,624,199,523,191,248,389,157,1196,498,455,210,1134,1246,364,750,685,1123,535,446,568,213,1336,1114,1587,370,1002,884,1536,255,205,1484,1121,1597,1624,1007,891,1468,444,890,1502,577,1519,1376,1549,587,1544,616,1014,640,1574,916,1566,824,926,588,953,986,1571,1501,335,330,897,843,1577,1374,917,1348,316,1417,281,1162,838,551,525,875,691,1540,1359,870,1170,1176,793,1369,1285,701,593,1398,1321,1024,942,1555,1560,1629,1276,1505,1503,870,486,1170,1201,902,350,166,338,1093,464,494,399,889,842,481,93)

>K01<-
c(225,452,505,624,199,248,191,535,568,1114,370,890,587,824,1374,316,1276,1555,166,1093)

> length(M401)

> intersect(kopyacılar,M401)

> length(intersect(kopyacılar,M401))

> length(K01)

```
> intersect(kopyacılar,K01)

> length(intersect(kopyacılar,K01))

>devriyeler<-union(lzstarcıktı,CUSUMcıktı)

>devriyeler<-union(devriyeler,ZU3cıktı)

>Model1<-union(devriyeler,LNMRT)

>Model1<-union(Model1,K01)

>length(intersect(Model1,kopyacılar)

>Model2<-union(devriyeler,K01)

>Model2<-union(Model2,ln)

>length(intersect(Model2,kopyacılar)

>Model3<-union(devriyeler,LNMRT)

>Model3<-union(Model3,M401)

>length(intersect(Model3,kopyacılar)

>Model4<-union(devriyeler,ln)

>Model4<-union(Model4,M401)

>length(intersect(Model4,kopyacılar)
```

#Denetimli Sınıflandırma Yöntemleri İçin R Kodları

##Türetilmiş değişkenler;

```
library(irtoys)

library(lattice)
```

```
library(LNIRT)

library(ltm)

##Kişi yetenelekleri;

yetenek<-ability(yanıtlarikili, method = "WLE", mu = 0, sigma = 1, n = 5)

##Kişi hızları;

out=LNIRT(RT=RT,Y=yanıtlarikili,WL=1,XG=nrep)

alpha=out$Post.Means$Sigma2 #Zaman ayrımcılığı parametresi

beta=out$Post.Means$Time.Intensity #Zaman yoğunluğu parametresi

hız=function(alpha,beta,RT)

{return(rep(sum(alpha*alpha*beta)/sum(alpha*alpha),nmen)-

        RT%*%(alpha*alpha)/(sum(alpha*alpha)))}

##Girdi değişkenleri;

> yanıtlar<-read.table("yanıtlarikili.txt",header=T)

> rt<-read.table("RT.txt",header=T)

> demografik<-read.table("demografik.txt",header=T)

> tsüre<-demografik[,10]

>basari<-read.table(“basari.txt”,header=T)

> hız<-read.table("hız.txt",header=T)

> yetenek<-read.table("yetenek.txt",header=T)

> lzstar<-read.table("lzstarcıktı.txt",header=T)

> ZU3<-read.table("ZU3cıktı.txt",header=T)
```

```
>cusumartı<-read.table(“cartı.txt”,header=T)

>cusumeksi<-read.table(“ceksi.txt”,header=T)

> lt<-read.table("ln.txt",header=T)

> flagged<-demografik[,3]

> veri<-cbind(yanıtlarikili,rt)

> veri<-cbind(veri,tsüre)

> veri<-cbind(veri,basari)

> veri<-cbind(veri,hız)

> veri<-cbind(veri,yetenek)

> veri<-cbind(veri,lzstar)

> veri<-cbind(veri,ZU3)

> veri<-cbind(veri,cusumartı)

> veri<-cbind(veri,cusumeksi)

> veri<-cbind(veri,lt)

> veri<-cbind(veri,flagged)

> veri<-edit(veri)
```

##Ölçekleme ve özellik seçimi;

```
> ölcekleme<-function(x,na.rm=T){
+     return((x-min(x))/(max(x)-min(x)))
+ }
> library(caret)
```

```
> veri$flagged = factor(veri$flagged, levels = c(0, 1))
> veri[-350]<-ölcekleme(veri[-350])
> control <- rfeControl(functions=rfFuncs, method="cv", number=10)
> results <- rfe(veri[,1:349], veri$flagged, sizes=c(1:349), rfeControl=control)
> predictors(results)
```

##Yeni veri setinin çağırılması;

```
veri<-read.table("yeniveri.txt",header=T)
```

#Yeni veri setinin eğitim veri seti ve test veri seti olarak ayrılması;

```
> ölcekleme<-function(x,na.rm=T){
+ return((x-min(x))/(max(x)-min(x)))
+ }
> veri$flagged = factor(veri$flagged, levels = c(0, 1))
> veri[-13]<-ölcekleme(veri[-13])
> library(caTools)
> set.seed(123)
> split = sample.split(veri$flagged, SplitRatio = 0.70)
> egitimset = subset(veri, split == TRUE)
> testset = subset(veri, split == FALSE)
```

##Denetimli sınıflandırma yöntemleri;

```
> library(randomForest)
> classifier = randomForest(x = egitimset[-13],
```

```
+ y = egitimset$flagged,
+ ntree = 1000)
> y_pred = predict(classifier, newdata = testset[-13])
> cm = table(testset[, 13], y_pred)
> cm
> library(e1071)
> classifier = naiveBayes(x = egitimset[-13],
+ y = egitimset$flagged)
> y_pred = predict(classifier, newdata = testset[-13])
> cm = table(testset[, 13], y_pred)
> cm
> library(class)
> y_pred = knn(train = egitimset[,13],
+ test = testset[,13],
+ cl = egitimset[,13],
+ k = 5)
> cm = table(testset[, 13], y_pred)
> cm
```

İlgilenen okuyucular çalışma veri seti ve gerçekleştirilen analizler için R Programı kodlarına ve çıktılarına https://github.com/muhammetkumartas/kulyutmaz üzerinden erişebilirler.

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Muhammet Kumartaş

Eğitim Durumu

Lisans Öğrenimi : Necmettin Erbakan Üniversitesi, Ahmet Keleşoğlu Eğitim Fakültesi, Rehberlik Ve Psikolojik Danışmanlık Bölümü.

Yüksek Lisans Öğrenimi : Akdeniz Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme ve Değerlendirme Tezli Yüksek Lisans Programı.

Bildiği Yabancı Diller : İngilizce.

Tarih : 12/07/2023

BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kâğıt ve elektronik kopyalarının Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

☐ Tezimin/Raporumun tamamı her yerden erişime açılabilir.

☐ Tezim/Raporum sadece Akdeniz Üniversitesi yerleşkelerinden erişime açılabilir. Tezimin/Raporumun 3 yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

12/07/2023

Muhammet Kumartaş

İNTİHAL RAPORU



Doküman Görüntüleyici

Turnitin Orijinallik Raporu

İşleme konusu: 10-Ağu-2023 12:49 +03
NUMARA: 2143905783
Kelime Sayısı: 24358
Gönderildi: 1

ÇEVRİMİÇİ SINAVLARDA TEST HİLESİNİN BELİRLENM... Muhammet Kumartaş tarafından

Benzerlik Endeksi %12

Kaynağa göre Benzerlik
Internet Sources: %11
Yayınlar: %2
Öğrenci Ödevleri: %3

alıntıları çıkar | bibliyografyayı dahil et | küçük eşleşmeleri çıkar | mod: raporu hızlı görüntüle (klasik) | yazdır | yenile | İndir

2% match (26-May-2023 tarihli internet)
https://github.com/muhammetkumartas/kulyutmaz/blob/main/PFSS.txt

1% match (11-Haz-2023 tarihli internet)
https://github.com/muhammetkumartas/kulyutmaz/blob/main/It(%C4%B1rt).txt

1% match (24-Haz-2023 tarihli internet)
https://github.com/muhammetkumartas/kulyutmaz/blob/main/veri4SONUCLAR.txt

1% match (11-Haz-2023 tarihli internet)
https://github.com/muhammetkumartas/kulyutmaz/blob/main/kodlar.txt

<1% match (24-Haz-2023 tarihli internet)
https://github.com/muhammetkumartas/kulyutmaz/blob/main/veri1SONUCLAR.txt

<1% match (11-Haz-2023 tarihli internet)
https://github.com/muhammetkumartas/kulyutmaz/blob/main/notlar.txt

<1% match (24-Haz-2023 tarihli internet)
https://github.com/muhammetkumartas/kulyutmaz/blob/main/veri3SONUCLAR.txt

<1% match (24-Haz-2023 tarihli internet)
https://github.com/muhammetkumartas/kulyutmaz/blob/main/veri2SONUCLAR.txt

<1% match (26-May-2023 tarihli internet)
https://github.com/muhammetkumartas/kulyutmaz/blob/main/RAPOR.txt

<1% match (02-Oca-2017 tarihli öğrenci ödevleri)
Submitted to Akdeniz University on 2017-01-02

<1% match (03-Haz-2021 tarihli öğrenci ödevleri)
Submitted to Akdeniz University on 2021-06-03

<1% match (06-Haz-2023 tarihli öğrenci ödevleri)
Submitted to Akdeniz University on 2023-06-06

<1% match (20-Kas-2019 tarihli öğrenci ödevleri)
Submitted to Akdeniz University on 2019-11-20