



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

KARAR AĞACI ALGORİTMASININ BİLİŞSEL TANI MODELLERİNE DAYALI
BİLGİSAYARDA BİREYSELLEŞTİRİLMİŞ TESTLERDE KULLANIMI

Hüseyin YILDIZ

Doktora

Ankara, 2024



Liderlik, araştırma, inovasyon, kaliteli eğitim ve deęişim ile

Daha ileriye... En İyeye...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

KARAR AĞACI ALGORİTMASININ BİLİŞSEL TANI MODELLERİNE DAYALI
BİLGİSAYARDA BİREYSELLEŞTİRİLMİŞ TESTLERDE KULLANIMI

THE USE OF DECISION TREE ALGORITHM IN COMPUTERIZED ADAPTIVE TESTS
BASED ON COGNITIVE DIAGNOSTIC MODELS

Hüseyin YILDIZ

Doktora

Ankara, 2024

Kabul ve Onay

Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Hüseyin YILDIZ'ın hazırladığı “Karar Ağacı Algoritmasının Bilişsel Tanı Modellerine Dayalı Bilgisayarda Bireyselleştirilmiş Testlerde Kullanımı” başlıklı bu çalışma jürimiz tarafından **Eğitim Bilimleri Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiştir.

Jüri Başkanı	Prof. Dr. Selahattin GELBAL	İmza
Jüri Üyesi (Danışman)	Prof. Dr. Hülya KELECİOĞLU	İmza
Jüri Üyesi	Prof. Dr. Neşe GÜLER	İmza
Jüri Üyesi	Prof. Dr. Dilara BAKAN KALAYCIOĞLU	İmza
Jüri Üyesi	Doç. Dr. Kübra ATALAY KABASAKAL	İmza

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim, Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından 11/ 07/ 2024 tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca / / tarihi itibarıyla kabul edilmiştir.

Prof. Dr. İsmail Hakkı MİRİCİ
Eğitim Bilimleri Enstitüsü Müdürü

Öz

Bilişsel tanı modelleri (BTM) ve bilgisayar ortamında bireye uyarlanmış testler (BBT) eğitimde ölçme ve değerlendirme alanında halen gelişmekte olan ve alan yazında popülerliğini koruyan konulardandır. Bu çalışmada Bilişsel Tanı Modellerine dayalı bireyselleştirilmiş testler için (BT-BBT), karar ağacı algoritmalarına dayalı başlangıç, ilerleme ve sonlandırma kurallarını kapsayan bir yöntem önerilmiştir. Önerilen yöntemi geleneksel BT-BBT uygulamaları ile farklı madde havuzu büyüklüğü, örneklem büyüklüğü ve test uzunluğu koşulları altında örtük sınıf ve beceri kestirim tutarlılığı açısından karşılaştıran simülasyon verileri ile yürütülmüştür. Ayrıca gerçek öğrenci tepkileri kullanılarak da benzer karşılaştırmalar yapılmıştır. Araştırmada kullanılan verilerin türetilmesi ve analizler R programlama dilinde çeşitli paket programlar kullanılarak gerçekleştirilmiştir. Araştırmada gerçek veri kullanılarak elde edilen sonuçlara göre, karar ağacına dayalı yöntem kullanılarak elde edilen başarılı sınıf ve beceri tahmini oranlarının geleneksel yöntemlere göre daha yüksek olduğu gözlenmiştir. Ayrıca 18 ve 24 madde uzunluklarının kullanıldığı simülasyon koşullarında karar ağacına dayalı yöntem daha başarılı olurken 12 maddelik test uzunluklarının kullanıldığı koşullarda Global Discrimination Index (GDI) ilerleme kuralının kullanıldığı geleneksel uygulamalar daha yüksek başarılı sınıf tahmini oranları elde etmiştir.

Anahtar sözcükler: bilişsel tanı modelleri, bilgisayarda bireyselleştirilmiş test, G-DINA, karar ağacı, makine öğrenmesi

Abstract

Cognitive Diagnostic Models (CDM) and Computerized Adaptive Testing (CAT) remain topics of ongoing development and popularity in the field of educational measurement. This study proposes a method for adaptive tests based on Cognitive Diagnosis Models (CDM-CAT) that incorporates starting, item selection, and termination rules grounded in decision tree algorithms. Simulations were conducted to compare the proposed method with traditional CDM-CAT applications under varying conditions of item pool size, item discrimination levels, and test length, assessing accuracy of latent class and skill estimation. Additionally, similar comparisons were made using real student responses. Data derivation and analysis were carried out using various packages in the R programming language. Results show that the decision tree-based method achieved higher rates of successful class and skill estimation compared to traditional methods when using real data. Moreover, under simulation conditions with test lengths of 18 and 24 items, the decision tree-based method demonstrated better performance. In contrast, traditional applications employing the Global Discrimination Index (GDI) progression rule yielded higher successful classification rates in conditions utilizing 12-item test lengths.

Keywords: cognitive diagnostic models, computerized adaptive testing, decision tree, machine learning

Teşekkür

Doktora ders döneminden itibaren bana kattıkları, tez yazım sürecimdeki kıymetli katkıları ve rehberliği için danışmanım Prof. Dr. Hülya KELECİOĞLU'na şükranlarımı sunarım. Tez önerimi sunduğum ilk günden bu yana değerli görüş ve katkılarından dolayı Prof. Dr. Neşe GÜLER'e ve Doç. Dr. Kübra ATALAY KABASAKAL'a teşekkür ederim. Yüksek lisans ve doktora eğitimim boyunca Eğitimde Ölçme ve Değerlendirme alanındaki gelişimime yardımcı olan tüm hocalarıma teşekkürü borç bilirim.

Ayrıca bu süreçte hem alan bilgileriyle hem de dostluklarıyla yanımda olan başta Dr. Alperen YANDI, Dr. Erdem BODUROĞLU, Dr. Mahmut Sami YİĞİTER olmak üzere tüm arkadaşlarıma teşekkür ederim.

Küçük yaşlardan beri eğitimin önemini bize anlatıp iyi bir eğitim almamız için maddi manevi her türlü desteği bize koşulsuz sağlayan dedem, babam, anneme ve ayrıca varlıklarından güç aldığım abim ve kardeşime minnettarım.

En özel teşekkürlerimi ise, bu tezin hazırlanmasında büyük payı olan, profesyonel kariyerimde beni desteklemek için en büyük fedakârlıkları tereddütsüz gösteren, yol arkadaşım, değerli meslektaşım ve eşim Dr. Eda AKDOĞDU YILDIZ'a sunuyorum.



Eda ve Ege'ye...

İçindekiler

Kabul ve Onay	ii
Öz	iii
Abstract.....	iv
Teşekkür	v
Tablolar Dizini	ix
Şekiller Dizini	x
Simgeler ve Kısaltmalar Dizini	xi
Bölüm 1 Giriş	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	2
Araştırma Problemi.....	3
Sayıtlılar.....	3
Tanımlar.....	4
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	5
Bilgisayarda Bireyselleştirilmiş Testler.....	5
Bilişsel Tanı Modelleri.....	10
Bilişsel Tanı Modellerine Dayalı Bilgisayarlı Bireyselleştirilmiş Testler (BT-BBT)	13
Karar Ağacı Algoritması.....	15
Karar Ağaçlarına Dayalı BT-BBT	18
İlgili Araştırmalar	21
Bölüm 3 Yöntem	24
Araştırmanın Türü.....	24
Araştırmanın Deseni.....	24
Verilerin Tanıtılması.....	27

Simülasyon Koşulları	29
Verilerin Analizi	33
Bölüm 4 Bulgular, Yorumlar ve Tartışma	36
1. Alt Probleme İlişkin Bulgular	36
2. Alt Probleme İlişkin Bulgular	50
Bölüm 5 Tartışma, Sonuç ve Öneriler	55
Araştırmacılara Öneriler	57
Uygulayıcılara Öneriler	57
Kaynaklar	59
EK-A: Araştırma Etik Komisyon İzin Muafiyeti Formu	64
EK-B: Etik Beyanı	65
EK-C: Doktora Tez Çalışması Orijinallik Raporu	66
EK-Ç: Dissertation Originality Report	67
EK-D: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	68

Tablolar Dizini

Tablo 1 <i>Q Matris</i>	28
Tablo 2 <i>Madde Parametrelerinin Üretim Şartları</i>	31
Tablo 3 <i>Değişimlenen Simülasyon Koşulları</i>	33
Tablo 4 <i>Farklı Simülasyon Koşullarına Göre Başarılı Örtük Sınıf Tahmini Oranları</i>	37
Tablo 5 <i>Farklı Simülasyon Koşullarına Göre Başarılı Beceri Tahmini Oranları</i>	39
Tablo 6 <i>Farklı Madde Ayırıcılık Düzeylerine Göre Başarılı Sınıf Tahmini Oranları</i>	40
Tablo 7 <i>Farklı Madde Ayırıcılık Düzeylerine Göre Başarılı Beceri Tahmini Oranları</i>	42
Tablo 8 <i>Farklı Test Uzunluklarına Göre Başarılı Sınıf Tahmini Oranları</i>	43
Tablo 9 <i>Farklı Test Uzunluklarına Göre Başarılı Beceri Tahmini Oranları</i>	45
Tablo 10 <i>Farklı Madde Havuzu Büyüklüklerine Göre Başarılı Sınıf Tahmini Oranları</i>	47
Tablo 11 <i>Farklı Madde Havuzu Büyüklüklerine Göre Başarılı Beceri Tahmini Oranları</i>	49
Tablo 12 <i>Gerçek Veri Kullanılarak Elde Edilen Başarılı Sınıf ve Beceri Tahmini Oranları</i>	51
Tablo 13 <i>Karar Ağacı Uygulamasında Test Uzunluklarına Ait Frekans ve Yüzde Değerleri</i>	52
Tablo 14 <i>Karar Ağacı ve GDI Uygulamalarından Elde Edilen Maruz Kalma Oranları</i>	53

Şekiller Dizini

Şekil 1 <i>BBT Uygulamalarının Akış Şeması</i>	5
Şekil 2 <i>Karar Ağacına Dayalı Yöntemin Akış Şeması</i>	21
Şekil 3 <i>Araştırma Deseni Şeması</i>	25
Şekil 4 <i>Ayırıcılık Düzeylerine Göre Başarılı Sınıf Tahmini Oranları Grafiği</i>	41
Şekil 5 <i>Ayırıcılık Düzeylerine Göre Başarılı Beceri Tahmini Oranları Grafiği</i>	42
Şekil 6 <i>Farklı Test Uzunluğu Koşullarına Göre Başarılı Sınıf Tahmini Oranları Grafiği</i>	44
Şekil 7 <i>Farklı Test Uzunluğu Koşullarına Göre Başarılı Beceri Tahmini Oranları Grafiği</i>	46
Şekil 8 <i>Madde Havuzu Büyüklüğüne Göre Başarılı Sınıf Tahmini Oranları</i>	48
Şekil 9 <i>Madde Havuzu Büyüklüğüne Göre Başarılı Beceri Tahmini Oranları</i>	49

Simgeler ve Kısaltmalar Dizini

BBT: Bilgisayarlı Bireyselleştirilmiş Testler

BTM: Bilişsel Tanı Modelleri

BSTO: Başarılı Sınıf Tahmini Oranı

BBTO: Başarılı Beceri Tahmini Oranı

PWKL: Posterior Weighted Kullback Leibler

BT-BBT/ CD-CAT: Bilişsel Tanı Modellerine Dayalı Bireye Uyarlanmış Testler

G-DINA: Generalized Deterministic Inputs, Noisy “And” Gate

GDI: Global Discrimination Index

BEDSV: Beklenen En Düşük Sonsal Varyans

MLE: Maximum Likelihood Estimation/ Maksimum Olabilirlik

EAP: Expected a Posteriori/ Beklenen Sonsal

DKL: Double Kullback Leibler

KA-2: 2 maddelik öbeklerin kullanıldığı karar ağacı modelleri

KA-3: 3 maddelik öbeklerin kullanıldığı karar ağacı modelleri

KA-4: 4 maddelik öbeklerin kullanıldığı karar ağacı modelleri

Bölüm 1

Giriş

Bu bölümde problem durumu, araştırmanın amacı, problem cümlesi ve alt problemlere yer verilmiştir.

Problem Durumu

Eğitimde ve psikolojide Bilgisayarda Bireyselleştirilmiş Testlerin (BBT) kullanımı her geçen gün artmakta ve yeni yöntemsel çalışmalar yayınlanmaktadır. BBT uygulamalarının daha az madde kullanımıyla daha kesin ölçmelere imkân vermesi bu araçları cazip hale getirmektedir (Weiss & Vale, 1987). Öte yandan öğrenme eksikliklerini daha ayrıntılı incelemeye imkân veren, bireyler hakkında sadece sonuç değil aynı zamanda süreç odaklı bilgiler edinmeye katkı sağlayan Bilişsel Tanı Modelleri (BTM) de üzerinde artan bir sıklıkla çalışılan bir başka konudur.

Bu iki işlevsel araç 1980'li yıllarda yapılan çalışmalarla birleştirilip, BTM'ye dayalı sınıflama işlemleri BBT aracılığıyla uygulanmaya başlanmıştır (Tatsuoka, 1983; Tatsuoka & Tatsuoka, 1997; Tatsuoka & Tatsuoka, 1989). İlk uygulamaların yapıldığı dönemden beri bilişsel tanı modellerine dayalı bireyselleştirilmiş bilgisayarlı testler (BT-BBT) ile ilgili birçok çalışma yapılmış ve farklı teorik yaklaşımlar önerilmiştir (McGlohen & Chang, 2008; Cheng, 2009; Wang, 2013; Wang ve ark., 2020).

BT-BBT uygulamalarından elde edilen sınıflama kararlarının tutarlılığını etkileyen birçok değişken olduğu söylenebilir (Cheng, 2010). Bu değişkenlerden bazıları, madde seçim (ilerleme) yöntemleri, madde havuzunun büyüklüğü, Q matrisinde tanımlanmış nitelik sayısı ve sonlandırma kuralları olarak sıralanabilir. Söz konusu değişkenlerin tek başına ve birlikte değişimleri BT-BBT uygulamalarının başarılı sınıf kestirimi oranlarına doğrudan etki etmektedir. Bu sebeple alan yazında yapılan birçok çalışma ile sınıflama tutarlılığını arttıracak ideal koşul ve yöntemler tespit edilmeye çalışılmıştır (Chen ve ark., 2012; Guo &

Zeng, 2019; Liu ve ark., 2016; Madison & Bradshaw, 2015; Sorrel ve ark., 2020; Tatsuoka, 2002).

Geleneksel BT-BBT uygulamaları karmaşık matematiksel altyapısı sebebi ile araştırmacıların ve uygulayıcıların gerçek uygulamalar yapmasına imkan sağlayacak platform yazılımlarına sahip değildir. Yazılım geliştiriciler halihazırda aktif kullanıcı sayısının fazla olmaması sebebi ile uygun bir platform geliştirmezken, araştırmacı ve uygulayıcılar için ise, karmaşık matematik ve önceden planlanmış bir ilerleme örüntüsünün olmayışı kendi uygulamaları için bir platformu geliştirmelerini güçleştirmektedir. Yazılım konusundaki bu engellerin, yöntemin yaygınlaşmasının önünde önemli bir engel olduğu söylenebilir.

Araştırmanın Amacı ve Önemi

Bu araştırmanın amacı bilişsel tanı modellerine dayalı bireye uyarlanmış testlerde karar ağacı algoritmalarına dayalı bir madde seçim yöntemi önermek ve önerilen yöntemi çeşitli koşullar altında Posterior Weighted Kullback Leibler (PWKL) ve G-DINA (Generalized Deterministic Inputs, Noisy “And” Gate), Global Discrimination Index (GDI) yöntemleri ile sınıflama başarısı açısından karşılaştırmaktır.

BT-BBT uygulamalarında kullanılan yöntem ve koşulların sınıflama tutarlılığı ve test uzunluklarına etkisi bir Monte Carlo simülasyon süreci ile incelenerek alan yazına katkı sağlamayı amaçlamaktadır. Ayrıca araştırmada karar ağacı algoritmaları kullanarak yenilikçi bir BT-BBT uygulama modeli önerilmektedir. Söz konusu yöntem önerisinin amacı BBT uygulamalarının daha az madde kullanılarak ya da daha düşük hata oranları ile gerçekleştirilebilmesine katkı sunmaktır. Çalışmanın bir diğer amacı ise karar ağacına dayalı önerilen yöntemin ilerleme örüntüsünün kalibrasyon aşamasında belirlenmesinden dolayı, geleneksel BT-BBT uygulamalarının hesaplama yazılım geliştirme gibi dezavantajlarını ortadan kaldırmaya katkı sağlamasıdır.

Araştırma Problemi

Bilişsel tanı modellerine dayalı bilgisayarda bireyselleştirilmiş testlerde karar ağacı algoritmasına dayalı madde seçim yöntemi ile Posterior Wiegthed Kulback Liebler (PWKL) ve G-DINA Discrimination Index (GDI) yöntemlerinin sınıflama tutarlılıkları test uzunluklarına, madde ayıricılık gücü düzeylerine ve madde havuzu büyüklüğüne göre farklılaşmakta mıdır?

Alt Problemler

1. Simülasyon verisinde karar ağacına dayalı uygulama ve BT-BBT uygulamaları ile elde edilen başarılı beceri ve sınıf tahmini oranları,
 - 1.1. madde ayıricılık düzeylerine göre nasıldır?
 - 1.2. farklı test uzunluklarına göre nasıldır?
 - 1.3. farklı madde havuz büyüklüklerine göre nasıldır?
2. Gerçek veriye dayalı simülasyonla yapılan karar ağacı ve BT-BBT uygulamalarında farklı yöntemlere göre (Global Discrimination Index ve Posterior Weighted Kullback Liebler),
 - 2.1. başarılı beceri ve sınıf tahmini oranları nasıldır?
 - 2.2. madde maruz kalma oranları nasıldır?

Sayıtlar

Araştırmada kullanılan verilerin bir kısmı literatüre dayalı belirli parametreler baz alınarak yapay olarak üretilmiştir. Yapay verilerin gerçek öğrenci tepkilerini kabul edilebilir ölçüde yansıttığı varsayılmıştır. Ayrıca araştırmada gerçek öğrenci tepkilerinden elde edilen madde parametreleri ve gerçek maddeler incelenerek oluşturulan Q matrisleri kullanılmıştır. Bu parametreler ve Q matrisine dayanarak öğrencilerin bireyselleştirilmiş testlere vereceği tepkiler yapay olarak oluşturulmuştur ve gerçek tepkşlerşni yansıtacağı varsayılmıştır.

Tanımlar

Araştırmada madde öbekleri kavramı, karar ağacının her bir düğümünde yer alan madde grubuna verilen isimdir.



Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde araştırma konusuna ait kuramsal çerçeveye ait bilgiler sunulmuştur.

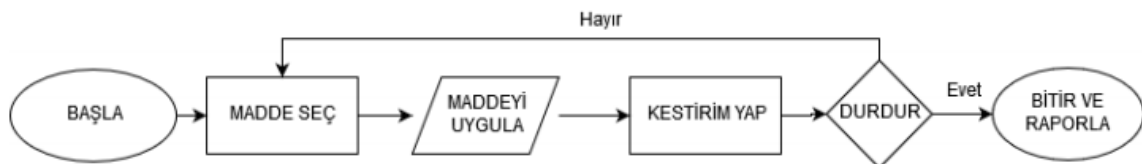
Bilgisayarda Bireyselleştirilmiş Testler

Bilgisayarda Bireyselleştirilmiş Testler (BBT) son dönemde sıklıkla kullanılan ve alan yazında ilgili teorik ve pratik çalışmaların artarak devam ettiği önemli bir araçtır. BBT'nin temel mantığı bireye ait örtük yetenek düzeyinin test sonunda bir defa kestirilmesi yerine, bireye yönlendirilen her maddeden sonra yetenek düzeyini tekrar tekrar kestirip, çeşitli yöntemlerle kestirilen yetenek düzeyine uygun maddenin yönlendirilmesi prensibine dayanmaktadır (Bartram & Hambleton, 2006; de Ayala, 2009). Bu sayede bireyin yetenek düzeyine uygun olmayan maddelerle karşılaşmasına engel olarak test uzunluklarını kısaltmak mümkündür. BBT uygulamalarında gruptaki bireylere uygulanan maddeler ve test uzunlukları farklılık gösterebilmektedir. Bu yöntemin bir diğer avantajı da ölçme sonuçlarının uygulama bitiminde anlık olarak gözlemlenebilmesidir.

Tüm bu avantajlarının yanında BBT uygulamalarının önceden kalibre edilmiş geniş madde havuzları gerektirmesi ve test uzunluğunun azaltılması karşılığında bir miktar standart hatanın kabul edilmesi gibi dezavantajları olabilir. BBT uygulamalarının çalışma prensibi Şekil 1'de şema ile gösterilmiştir (Davey, 2011).

Şekil 1

BBT Uygulamalarının Akış Şeması



Şekil 1'de görüldüğü gibi her BBT uygulaması önceden tanımlanmış bir kurala göre seçilen bir madde ile başlar. Bireyin vereceği cevaba göre yetenek kestirimi yapılır. Kestirim

sonrası yine önceden belirlenmiş test sonlandırma kuralının gerçekleşip gerçekleşmediği kontrol edilir. Eğer sonlandırma kuralı gerçekleştiyse test sonlandırılır ve nihai yetenek kestirimi raporlanır. Sonlandırma kuralı gerçekleşmediyse belirlenen madde seçim kuralına uygun şekilde madde havuzundan yeni bir madde seçilip bireye yönlendirilir ve döngü devam eder. Tüm bu aşamalar başlangıç kuralı, ilerleme kuralı, sonlandırma kuralı olarak 3 başlık altında değerlendirilebilir.

Başlangıç Kuralı

BBT uygulamaları yetenek düzeylerine uygun maddelerin yönlendirilmesi esasına dayalıdır. Katılımcılara ait ölçülen özellik bakımından düzeylerini temsil eden önsel puanlar ya da yetenek düzeyleri biliniyorsa, her bir öğrencinin testi bu yetenek düzeyinde en yüksek bilgi üreten madde ile başlatılabilir. Fakat çoğu zaman uygulama öncesi bireylerin yetenek düzeyleri ile ilgili önsel bir bilgiye sahip olunmayabilir. Yaygın olarak ortalama yetenek düzeyinde yüksek bilgi üreten maddelerle teste başlanması tercih edilir. MTK'ye dayalı yetenek düzeyi ölçeklendirmesinde “theta=0” ortalama yetenek düzeyi olarak kabul edilir. Bu sebeple sıklıkla “theta=0” düzeyinde en yüksek bilgi üreten madde ile başlanabilmektedir. Fakat madde havuzunda bu düzeyde en yüksek bilgiyi üreten tek bir madde olduğunda söz konusu madde grubun tamamına uygulanmış olacaktır. Bu durum test güvenliği açısından (madde ifşası) sakıncalı olabilir. Ayrıca tüm gruba ilk madde olarak ortak maddenin yönlendirilmesi ilk madde sonrası doğru ve yanlış yanıt verenlere göre sadece iki farklı yetenek kestiriminin ortaya çıkmasına sebep olacaktır. Dolayısıyla ikinci maddelerin sonuna gelindiğinde havuzdan toplam sadece 3 madde kullanılmış olacaktır. Bu durum da hem test güvenliğinin tehlikeye atılmasına hem de yetenek düzeylerinin belirli noktalara kadar farklılaştırılmamasına yol açabilir. Bu sorunların ortadan kaldırılmasına yönelik bir başka başlangıç kuralı ise “theta=0” gibi tek bir önsel yetenek varsayımı yerine (-1, +1) gibi önceden belirlenmiş yetenek aralığından her birey için ayrı ayrı rastgele seçilecek bir yetenek düzeyi kullanılabilir. Bu sayede bireylere yönlendirilecek maddeler ilk adımda farklılaşacağından test güvenliği ihlali olasılığı azaltılacaktır.

İlerleme Kuralı

Bireylere yönlendirilen ilk madde dışındaki maddelerin seçimi, madde havuzu kısıtlama uygulamaları, her yanıttan sonraki yetenek kestirim süreci ilerleme kuralı başlığı altında incelenebilir. Geleneksel bireyselleştirilmiş test uygulamalarında madde seçim kuralı için geliştirilmiş birçok yöntem olmakla birlikte en sık kullanılan yöntemin Fischer'ın Maksimum Bilgi Yöntemi olduğu söylenebilir. Bu yöntem madde havuzundan, bireyin kestirilen son yetenek düzeyinde en yüksek bilgiyi üreten maddenin bireye yönlendirilmesi prensibine dayalıdır. Maksimum bilgi yöntemi, yetenek düzeyinin sürekli bir değişken olmasını gerektirir. Bu gereksinim, maksimum bilgi yönteminin bilişsel tanı modellerine dayalı bireyselleştirilmiş testlerde kullanılmasını engeller (Sorrel, 2020). BT-BBT uygulamalarında kullanılmak için tasarlanmış ve en yaygın olarak kullanılan ilerleme yöntemleri, Kullback-Liebler (KL), Posterior Weighted Kullback-Liebler (PWKL), G-DINA Global Discrimination Index (GDI) gibi yöntemlerdir.

İlerleme kuralı çerçevesinde değerlendirilebilecek bir diğer kavram ise madde havuzu kısıtlamalarıdır. Bu kısıtlamalar her BBT uygulamasında kullanılsa da madde havuzu güvenliği ve madde havuzunun etkin kullanımı için tavsiye edilmektedir (Stocking, 1994). Bu kısıtlamalardan ilki maruz kalma oranı kontrolüdür (Item Exposure). Maruz kalma oranı, madde havuzundaki herhangi bir maddenin örneklemin maksimum ne kadarı tarafından kullanılabileceğini temsil eder. Bu oran kısıtlanarak özellikle havuzdaki yüksek ayırtıcılık değerine sahip olan maddelerin, yüksek bilgi üretmesinden kaynaklı olarak tüm bireylere yönlendirilmesinin önüne geçer. Böylece hem test güvenliğinin artırılmasına hem de madde havuzunun verimli kullanılmasına katkı sağlanmış olur. Bir diğer madde havuzu kısıtlama yöntemi ise kapsam dengelemedir. Bazı uygulamalarda madde havuzu maddelerin belirli özelliklerine göre gruplara ayrılmak istenebilir. Bu durumda kullanıcı bireylere yönlendirilen maddelerin söz konusu gruplardan belirli oranlarda seçilmesini tercih edebilir. Bu gibi durumlarda madde seçimi kapsam dengeleme kuralı ile kısıtlandırılabilir (Huang, 1996).

Test ilerleme koşulları altında değerlendirilebilecek son konu ise yönlendirilen her madde sonucunda yapılacak yetenek kestirim işleminde kullanılacak yöntemdir. Bu kapsamda Maksimum Olabilirlik (MLE), Beklenen Sonsal (EAP) gibi yetenek kestirim yöntemleri kullanıcı tarafından tercih edilebilir.

MLE yöntemi, gözlemlenen cevapların olasılığını en yüksek yapan yetenek düzeyini tahmin etmeye çalışır. MTK modellerinde, bireyin yetenek düzeyi (θ) ve madde parametreleri kullanılarak bireyin cevaplarının olasılıkları hesaplanır. MLE yöntemi, bu olasılıkların çarpımını maksimize eden yetenek düzeyini bulur (Embretson & Rase, 2000) ve Eşitlik 1'de yer almaktadır;

- θ : Bireyin yetenek düzeyi
- $P(X_i = 1|\theta)$: Bireyin i -inci maddeye doğru cevap verme olasılığı
- X_i : i -inci maddeye verilen cevap (1 doğru, 0 yanlış)
- $L(\theta)$: Olabilirlik fonksiyonu

$$L(\theta) = \prod_{i=1}^n P(X_i = 1|\theta)^{X_i} (1 - P(X_i = 1|\theta))^{1-X_i} \quad (1)$$

MLE tahmini, bu olabilirlik fonksiyonunu maksimize eden θ değerini bulur ve Eşitlik 2'de yer almaktadır;

$$\hat{\theta}_{MLE} = \operatorname{argmax}_{\theta} L(\theta) \quad (2)$$

Expepected a Posteriori (EAP) yöntemi ise, bireyin yetenek düzeyinin beklenen sonsal değerini tahmin eder. Bu yöntem, bireyin gözlemlenen cevapları ve yetenek düzeyine dair bir önsel dağılım kullanarak yetenek düzeyini tahmin eder. EAP yöntemi, Bayesci bir yaklaşımdır ve genellikle bireylerin yetenek dağılımının normal dağılım olduğunu varsayar (Embretson & Reise, 2000).

- θ : Bireyin yetenek düzeyi

- $P(\theta|X)$: Gözlemlenen cevaplar X verilmişken yetenek düzeyinin bir sonsal olasılığı
- $f(\theta)$: Yetenek düzeyinin önsel dağılımı
- $P(X|\theta)$: Yetenek düzeyi θ verilmişken gözlemlenen cevapların olasılığı ise,

sonsal olasılık Eşitlik 3'teki gibidir;

$$P(\theta|X) = \frac{P(X|\theta)f(\theta)}{\int P(X|\theta)f(\theta) d\theta} \quad (3)$$

EAP tahmini, sonsal olasılığın beklenen değeri Eşitlik 4'teki gibidir;

$$\hat{\theta}_{EAP} = \int \theta P(\theta|X) d\theta \quad (4)$$

Sonlandırma Kuralı

BBT uygulamalarında Şekil 1'de gösterilen döngü, önceden tanımlanmış bir sonlandırma kuralı gerçekleştiğinde tamamlanır. Bu kural kullanıcı tarafından istenildiği gibi tanımlanabilse de temel olarak iki tip sonlandırma kuralından bahsedilebilir. Bunlardan biri sabit test uzunluğuna dayalı sonlandırma kurallarıdır. Bu tip uygulamalarda her bireye eşit sayıda madde uygulanır ve bir bireye önceden belirlenmiş sayıda madde uygulandığında test sonlandırılır. Burada tavsiye edilen genel geçer bir test uzunluğundan söz edilemez. Kullanıcılar kabul edebilecekleri standart hata, madde havuzunun genişliği, havuzdaki maddelerin psikometrik özellikleri gibi kriterleri göz önünde bulundurarak çok çeşitli test uzunluklarını tercih edebilirler. Bir diğer sonlandırma kuralı ise standart hata üst sınırına dayalı sonlandırma kurallarıdır. Bu tip uygulamalarda her döngünün sonunda kestirilen yetenek düzeyine ait standart hatanın önceden belirlenmiş bir eşik değerin altına düştüğünde testin sonlandırılmasını sağlayabilirler. Bireylere uygulanan maddeler ve bu maddelere verdikleri yanıt örüntüleri farklılık gösterdiği için her birey için belirlenen standart hata sınırının altına düşmek için gerekli madde sayısı değişkenlik gösterebilir. Bir diğer deyişle bu tip sonlandırma kuralları kullanıldığında her birey için test uzunlukları farklılaşmaktadır.

Bazı bireyselleştirilmiş test uygulamalarında ise uygulayıcılar, sabit test uzunluğuna ve standart hata üst sınırına dayalı sonlandırma kurallarını birlikte kullanabilmektedir. Bu durumda uygulayıcı, test öncesi hem öğrencinin alabileceği madde sayısına ait bir üst sınır belirler hem de yetenek kestiriminin standart hatasına ilişkin bir eşik değeri belirler. Testi alan öğrenci, bu sonlandırma kurallarından herhangi birini gerçekleştirdiğinde test sonlanır.

Bilişsel Tanı Modelleri

Bilişsel Tanı Modelleri (BTM) bireylerin test maddelerine verdikleri tepkileri kullanarak, test maddelerinin ilişkili olduğu niteliklere sahip oluş durumlarını kestirip, bireyleri örtük sınıflara ayırmayı amaçlayan modellerdir. BTM'ler bireylere bir puan ya da yetenek düzeyi atama kaygısı taşımazlar. Amaç bireyin hangi niteliklere ne ölçüde sahip olduğunu belirleyerek öğrenme eksikliklerini ve öğrenme ihtiyaçlarını tespit etmektir (Cheng, 2009). Tek boyutlu MTK modellerinin aksine BTM'de testler birden fazla örtük özelliği ölçebileceği gibi, her bir madde de birden fazla özellik ile ilişkilendirilebilir. Söz konusu bu ilişkilendirme test geliştirme süreci öncesinde geliştirici tarafından Q matrisi adı verilen iki boyutlu bir matris ile tanımlanır.

Q matrisi satırlarda maddeleri, sütunlarda ise maddelerin ilişkili olabileceği nitelikleri barındırır. Testte yer alan i maddesi ile j niteliği arasındaki ilişki Q matrisinin i . satır, j . sütununda tanımlanır (Madison & Broadshaw, 2015). Eğer i maddesine doğru yanıt için j niteliği gerekliyse ilgili hücreye 1, gerekli değilse 0 girişi yapılır. Q matrisi BTM'nin en temel unsuru iken aynı zamanda en çok tartışılan ve eleştirilen kısımdır. Q matrisinin test geliştirici ya da uzmanların öznel yargılarına dayanması bu matrisin hazırlanışında yapılacak olası hatalara sebep olabilir. Q matrisin hazırlayıcısı bir maddenin bir j özelliğini ölçtüğünü düşünürken, madde j özelliği dışında başka özellikler kullanılarak doğru yanıtlanıyor olabilir. Bu gibi hatalar yapılan kestirim sonuçlarına hata karıştırabilir. Bu tip hataların mümkün olduğunca azaltılabilmesi için aşağıdaki önlemler alınabilir (Liu, Manley & Bradshaw, 2009).

- Q matrisini hazırlayan kişinin tanımlanmış beceriler ve bu becerilerin yapısı konusunda uzman olmalı.

- Test maddelerinin yazımından önce belirtke tablosuna dayanarak Q matrisi oluşturulmalı ve test maddeleri bu matris göz önünde bulundurularak yazılmalıdır.

- Matrisin hazırlanışında ve maddelerin beceriler ile ilişkilerinin ortaya konmasında birden fazla uzman görev almalı ve bu uzmanların oluşturdukları Q matrisler arasındaki uyum ortaya konmalı.

- Madde havuzunda Q matrisinde yer alan tüm becerileri yeterli düzeyde temsil edecek sayıda madde bulunmalı.

Q matrisi ne kadar hatadan arınık olursa yapılan kestirimler o derece hatasız ve model veri uyumu iyi olacaktır. BTM uygulamalarında ortaya çıkabilecek hata kaynakları da çoğunlukla Q matrisle ilgilidir.

Q matrisinde yer alan sütun sayısı, bir diğer deyişle test maddeleri ile ilişkilendirilen nitelik sayısı ortaya çıkacak örtük sınıf sayısı ile doğrudan ilişkilidir. k sayıda nitelik barındıran bir Q matrisi kullanan bir BTM analizi sonucunda bu özelliklere sahip olup olmama durumlarının tümünü barındıran 2^k sayıda örtük sınıf ortaya çıkacaktır. Uygulamaya katılan her birey bu 2^k örtük sınıftan birine yerleştirilir. Bu sınıf bireyin hangi beceriye sahip olup olmadığını tam olarak yansıtır.

Farklı teorik arka planlara sahip birçok bilişsel tanı modeli olmasına rağmen bu çalışmada yaygın olarak kullanılan modellerden GDINA (Generalized Deterministic Inputs, Noisy "And" gate) modeli kullanılmıştır. Aşağıda GDINA modelinin teorik arka planına ait bazı bilgiler sunulmuştur.

G-DINA Modeli

G-DINA modeli, bilişsel tanı modelleri arasında yaygın olarak kullanılan ve esnekliği ile öne çıkan bir modeldir. Bilişsel tanı modelleri, bireylerin belirli bilgi ve beceri setlerini ne kadar iyi öğrendiklerini ve bu setler arasındaki ilişkileri belirlemeyi amaçlar. GDINA modeli,

bu amaç doğrultusunda, bireylerin çeşitli görevlerdeki performanslarını analiz ederek, hangi beceri ve bilgi alanlarında eksiklikleri olduğunu belirler (de la Torre, 2011; Junker & Sijtsma, 2001; Rupp, Templin & Henson, 2010)

GDINA modeli, temel olarak her bir test maddesinin cevaplanması için gerekli olan bilgi ve beceri kümelerini tanımlar. Bu modelde, bir bireyin bir maddeyi doğru yanıtlatabilmesi için, o maddeyle ilişkili tüm gerekli bilgi ve becerilere sahip olması gerekir, ancak bu sahip olma durumu belirli bir hata payı (gürültü) ile değerlendirilir. Bu yaklaşım, GDINA modeline, madde yanıtlarının sadece gerekli bilgi ve becerilere bağlı olmadığı, aynı zamanda tesadüfi hataların da rol oynadığı gerçeğini dikkate alma esnekliği sağlar (de la Torre & Lee, 2013; DiBello, Roussos & Stout, 2007; Templin & Henson, 2006).

GDINA modelinin en önemli avantajlarından biri, çoklu becerilerin bir arada değerlendirilmesine imkan tanımasıdır. Diğer bilişsel tanı modellerine göre daha genel bir yapıya sahip olması, farklı beceri yapılarını ve maddelerin bu becerilerle ilişkisini daha doğru bir şekilde temsil etmesine olanak tanır. Bu esneklik, eğitimde ve psikometrik değerlendirmelerde GDINA modelinin yaygın olarak kullanılmasına katkıda bulunur. Ayrıca, modelin parametrik yapısı sayesinde, çeşitli madde türleri ve test formatlarıyla uyumlu bir şekilde kullanılabilir. Bu nedenle, GDINA modeli, eğitimsel değerlendirmeler ve bireysel öğrenme eksikliklerinin belirlenmesi için güçlü bir araç olarak kabul edilir (de la Torre & Minchen, 2014; Hartz & Roussos, 2008; Leighton & Gierl, 2007).

GDINA modeli, madde yanıt teorisi çerçevesinde, her bir maddenin çözülme olasılığını belirleyen çeşitli parametreleri kullanır. Bu bağlamda, özellikle s (slipping) ve g (guessing) parametreleri önemli rol oynar.

s (Slipping) Parametresi. s parametresi, bir bireyin gerekli tüm bilgi ve becerilere sahip olmasına rağmen bir maddeyi yanlış cevaplama olasılığını temsil eder. Bu, bireyin bildiği bir soruyu yanlışlıkla ya da dikkatsizlikten dolayı yanlış cevaplamasını ifade eder. S_i olarak ifade edilir. Burada i , maddeyi temsil eder. Dolayısıyla, s_i , i 'inci maddenin s parametresidir. Eğer bir birey gerekli bilgi ve becerilere sahipse (Q-matrix'e göre);

$$s_i = P(X_{ij} = 0 | \alpha_j = 1) \quad (5)$$

Eşitlik 5'teki gibi olur. Burada, X_{ij} , j'inci bireyin i'inci maddeye verdiği yanıtı, α_j , j'inci bireyin bilgi ve beceri düzeylerini temsil eder (Rupp, Templin & Henson, 2010).

g (Guessing) Parametresi. g parametresi, bir bireyin gerekli bilgi ve becerilere sahip olmamasına rağmen bir maddeyi doğru cevaplama olasılığını temsil eder. Bu, bireyin şans eseri doğru cevabı bulmasını ifade eder. g_i olarak ifade edilir. Yani, g_i , i'inci maddenin g parametresidir. Eğer bir birey gerekli bilgi ve becerilere sahip değilse;

$$g_i = P(X_{ij} = 1 | \alpha_j = 0) \quad (6)$$

Eşitlik 6'daki gibi olur. GDINA modelinde, bir bireyin i'inci maddeyi doğru cevaplama olasılığı, s ve g parametreleri kullanılarak Eşitlik 7'deki gibi ifade edilir;

$$P(X_{ij} = 1 | \alpha_j) = g_i + (1 - s_i - g_i) \cdot \prod_{k \in Q_i} \alpha_{jk} \quad (7)$$

- X_{ij} , j'inci bireyin i'inci maddeye verdiği yanıtı (1: doğru, 0: yanlış),
- α_j , j'inci bireyin bilgi ve beceri düzeylerini (latent traits),
- Q_i , i'inci madde için gerekli bilgi ve beceri kümelerini gösterir (Rupp, Templin & Henson, 2010).

Bilişsel Tanı Modellerine Dayalı Bilgisayarlı Bireyselleştirilmiş Testler (BT-BBT)

Klasik BBT uygulamaları çeşitli yöntemlerle mümkün olduğunca az madde kullanarak bireyin bir örtük özelliğe sahip oluş düzeyini kestirmeye çalışırlar. BT-BBT uygulamaları da benzer yöntemlerle mümkün olduğunca az madde kullanarak bireyin sahip olduğu nitelikleri ve ait oldukları örtük sınıfları kestirebilmeyi amaçlar. Yeteri kadar isabetli bir sınıflandırma yapmak için tanımlanan tüm niteliklere sahip olma durumlarının iyi tespit edilmesi gerekir.

BT-BBT uygulamalarında havuzdaki her madde, tüm niteliklerle ilişkili olmadığı için yönlendirilen her maddede tanımlı niteliklerin sadece bir kısmı ile ilgili bilgi edinilebilir. Örneğin başlangıç maddesi olarak i. madde yönlendirildiğinde bireyin yanıtı sonrası i. maddenin ilişkili olmadığı nitelikler ile ilgili bireyin yeterliği ile ilgili bir kestirim yapılamaz. Bu sebeple tanımlı tüm niteliklerle ilgili isabetli kararlar verebilmek için BT-BBT uygulamalarında klasik BBT'ye göre daha fazla sayıda madde kullanılması beklenebilir.

Bilişsel tanı modellerine dayalı bireyselleştirilmiş testlerde kullanılmak üzere geliştirilmiş birçok madde seçim yöntemi bulunmaktadır. Bu çalışmada literatürde yaygın olarak kullanılan Posterior Weighted Kullback-Leibler (PWKL) ve G-DINA discrimination index (GDI) yöntemleri tercih edilmiştir.

Posterior Weighted Kullback-Leibler (PWKL)

PWKL madde seçim yöntemi, bilişsel tanı modellerine (BTM) dayalı bireye uyarlanmış testlerde madde seçimi için kullanılan ileri bir tekniktir. PWKL yöntemi, bireyselleştirilmiş testlerde en uygun maddeyi seçmek için Kullback-Leibler (KL) bilgi ölçütünü kullanır. Bu yöntem, test edilen kişinin mevcut yanıtlarına dayalı olarak, her bir aday maddenin posterior olasılık dağılımlarını hesaplar (Kaplan ve ark., 2015). Ardından, bu dağılımlar kullanılarak her bir maddenin bilgi katkısı değerlendirilir.

$$PWKL_j(\hat{\alpha}_i^{(t)}) = \sum_{c=1}^{2^K} \left[\sum_{x=0}^1 \log \left(\frac{P(X_j = x | \hat{\alpha}_i^{(t)})}{P(X_j = x | \alpha_c)} \right) P(X_j = x | \hat{\alpha}_i^{(t)}) \pi_i^{(t)}(\alpha_c) \right] \quad (8)$$

Eşitlik 8'de $P(X_j = x | \alpha_c)$ ifadesi bireyin α_c beceri vektörüne sahip j maddesine x tepkisini vermesi olasılığını, $\pi_i^{(t)}(\alpha_c)$ ifadesi i 'inci bireyin t 'inci maddesine vereceği cevabın sonsal olasılığını temsil etmektedir.

PWKL yöntemi, yüksek bilgi katkısı sağlayan ve test edilen kişinin bilişsel profilini daha doğru bir şekilde belirlemeye yardımcı olan maddeleri seçmeyi hedefler. Böylece, testin genel doğruluğu ve etkinliği artırılırken, test edilen kişinin yanıtlarına dayalı olarak bireyselleştirilmiş bir şekilde ilerlenir (Cheng & Chang, 2009).

K-L bilgi düzeyi iki olasılık yoğunluk fonksiyonu $f(x)$ ve $g(x)$ arasındaki mesafenin bir ölçüsüdür; burada $f(x)$ 'in verilerin gerçek dağılımı olduğu varsayılır (Cover & Thomas, 1991).

G-DINA Discrimination Index (GDI)

GDI, bilişsel tanı modellerine (BTM) dayalı bireyselleştirilmiş testlerde madde seçimi için kullanılan bir ölçüttür ve de la Torre ve Chiu (2010) tarafından ortaya atılmıştır. GDI, test maddelerinin ayırt edicilik gücünü değerlendirmeye yardımcı olur ve test edilen bireylerin bilişsel profillerini daha doğru belirlemek için en uygun maddeleri seçmeyi amaçlar. Bu endeks, belirli bir gizli özellik dağılımı göz önüne alındığında, bir maddenin başarı olasılığının ağırlıklı varyansını hesaplar ve notasyonu Eşitlik 9'daki gibidir;

$$j = \operatorname{argmax}_{i \in B_q} GDI = E\{VAR[P_i(\theta)]\} = \int_{-\infty}^{\infty} P_i^2(\theta)W(\theta) - [\int_{-\infty}^{\infty} P_i(\theta)W(\theta)]^2 d\theta \quad (9)$$

Bu yöntem ağırlıklı doğru yanıt olasılığı varyansına dayalı olarak hesaplanan indekse göre (GDI) maddenin ayıricılık düzeyine karar verir. Bireyselleştirilmiş test algoritması indeks değeri en yüksek maddeyi seçer. Literatürde söz konusu indeks değeri için kabul gören bir eşik değeri olmamakla birlikte daha yüksek indeks değeri daha yüksek ayıricılık anlamına gelir (Shi ve ark., 2021).

Karar Ağacı Algoritması

Karar ağaçları makine öğrenmesi sınıflandırma algoritmaları arasında en yaygın kullanılanlardan birisidir. Bu teknikte bilinen verilerden yola çıkarak belirli kritik noktalarda dallara ayrılan bir ağaç oluşturulur. Daha sonra sınıflandırılmak istenen veri oluşturulan ağaca uygulanır. Ağacın her bir dalının sonunda bir sınıflandırma yargısı içeren bir yaprak bulunur. Silahtaroglu'na (2016) göre karar ağacı algoritmaları, ağacın oluşturulması ve her bir verinin ağaca uygulanarak sınıflandırılması olarak iki aşamadan oluşur. Bu algoritma karar ağacının oluşturulmasında entropi kavramını kullanır. Entropi bir veri grubundaki düzensizlik olarak tanımlanabilir (Dunham, 2003). Bir veri grubunun entropisi 0-1 aralığında

değer alır. Entropinin 0 olması veri grubunda hiç düzensizlik olmadığı anlamına gelir. Örneğin cinsiyetlerin tahmin edildiği bir veri grubunda herkesin kadın olması entropinin 0 olarak hesaplanmasına yol açar. Entropinin 1'e eşit olması veri grubundaki her bir bireyin tahmin edilmek istenen değişken açısından ayrı kategorilere ait olması durumu olarak düşünülebilir. Entropi değerini hesaplamak için Eşitlik 10'daki formül kullanılır.

$$H(S) = \sum_{c \in C} -p(c) \log_2 p(c) \quad (10)$$

$H(S)$ = S veri setine ait entropi değerini,

C = Tahmin edilecek değişkenin kategorileri,

$p(c)$ = S veri setinde C kategorisine ait veri oranını temsil eder.

Karar ağaçlarının oluşturulmasında kullanılan bir başka kavram bilgi kazanımı kavramıdır. S veri setinin dallara ayrılmadan önceki ve sonraki durumlardaki entropi miktarları arasındaki fark olarak tanımlanabilir. Karar ağaçlarından beklenen her bir dallanma sonra entropinin azalmasıdır. Karar ağaçları oluşturulurken önce ağacın kökü belirlenir. Bu kök ilk dallanmanın yaşandığı değişkendir. Birden fazla değişkenin söz konusu olduğu bir veri setinde hangi değişkenin ağacın kökünü oluşturulacağını belirlemek için, sırası ile her bir değişkene ait bilgi kazanımı değerleri hesaplanır. Bilgi kazanım değeri en yüksek olan değişken entropiyi yani düzensizliği en çok azaltacak değer olacağı için ağacın kökü olarak seçilir. Bilgi kazanımı değerini hesaplamak için kullanılan formül Eşitlik 11'de sunulmuştur.

$$IG(A, S) = H(S) - \sum_{t \in T} p(t) H(t) \quad (11)$$

$IG(A, S)$ = S veri setindeki A değişkenine ait bilgi kazanımı değerini,

t = A değişkeninin her bir alt kategorisini,

$p(t)$ = A değişkenindeki t kategorinde yer alan veri sayısının toplam veri sayısına oranını,

$H(t)$ = t kategorisine ait entropi değerini temsil etmektedir.

Veri setindeki her bir değere ait bilgi kazanımı değerleri hesaplandıktan sonra, en yüksek değere sahip değişken ağacın kökü olacaktır. Daha sonra köke ait her bir dal için aynı işlemler tekrar edilecektir.

Karar ağacı algoritmalarında aşırı öğrenme (overfitting), modelin eğitim verisine gereğinden fazla uyum sağlaması durumudur. Bu, modelin eğitim verisindeki tüm ayrıntıları ve gürültüyü öğrenmesiyle gerçekleşir. Aşırı öğrenme sonucunda, model eğitim verisi üzerinde çok yüksek bir doğruluk gösterirken, yeni ve görülmemiş veriler üzerinde kötü performans sergiler. Bu durum, modelin genelleme yeteneğinin zayıfladığı anlamına gelir, yani model, yeni verilerdeki desenleri ve ilişkileri doğru bir şekilde tahmin edemez (Quinlan, 1986).

Karar ağacı algoritmalarında budama (pruning), ağacın aşırı öğrenme riskini azaltmak için kullanılan bir tekniktir. Karar ağacı algoritmaları, veri kümesindeki desenleri öğrenmek için ağaç yapılarını kullanır. Ancak, ağaçlar bazı durumlarda eğitim veri kümesine çok fazla uyum sağlayabilir ve bu da genelleme performansını olumsuz etkileyebilir. Bu istenmeyen durumu engellemek için budama prosedürü uygulanabilir. Budama, ağacın aşırı öğrenmesini önlemek için ağacın gereksiz dallarını kaldırmak veya kısaltmak anlamına gelir (Quinlan, 1986). Bu, genellikle iki ana şekilde gerçekleştirilir:

Ön budama (Pre-pruning): Ön budama, ağaç oluşturulurken, ağacın büyümesini sınırlayan bir dizi kuralın uygulanmasıyla gerçekleşir. Bu kurallar, ağacın büyümesini kontrol eder ve aşırı öğrenmeyi engellemeye yardımcı olur. Örneğin, ağacın her bir dalına yerleştirilebilecek minimum gözlem sayısı için bir eşik değeri belirlenebilir. Bir diğer seçenek ise ağacın kökten itibaren maksimum kaç defa dallanabileceğini sınırlamaktır. Buna maksimum derinlik kuralı denir.

Sonradan budama (Post-pruning): Sonradan budama, ağaç oluşturulduktan sonra, gereksiz dalların kaldırılması veya kısaltılmasıyla gerçekleşir. Sonradan budama sırasında, her bir düğümün alt ağaçları, onları kaldırmak veya bir düğümü yalnızca bir yaprak

düğümüne dönüştürmek gibi belirli kriterlere göre değerlendirilir. Bu kriterler genellikle düğümün veri kümesine katkısını ve genelleme performansını dikkate alır.

Budama, karar ağacı modellerinin daha iyi genelleme yapmasına ve daha az karmaşık ve daha anlaşılır hale gelmesine yardımcı olabilir. Ayrıca, gereksiz dalların kaldırılması, modelin daha hızlı eğitilmesine ve daha hızlı tahmin yapmasına da yardımcı olabilir (Quinlan, 1986).

Karar ağaçlarının her bir yaprağındaki entropiyi 0'a indirene kadar dallanma işlemine devam etmesi, eğitim için kullanılan veri seti üzerinde aşırı öğrenmeye sebep olabilir. Bu tip durumlarda oluşturulan ağaç modeli eğitim verisi üzerinde çok başarılı görünürken, aynı evrene ait başka veri setlerine uygulandığında aynı başarıyı gösteremeyebilir. Bu sorunun önüne geçmek için karar ağaçlarının tüm dalları oluşturulduktan sonra budama işlemi uygulanır.

Karar Ağaçlarına Dayalı BT-BBT

Bilişsel Tanı Modelleri (BTM) puan ya da yetenek kestirimleri yapmak yerine, bireyleri sahip oldukları ya da sahip olmadıkları özellikler bakımından örtük sınıflara ayırmayı hedeflemektedir (Başokçu, 2014). Burada bahsedilen özellik kavramı; vasıf, görev, alt görev ya da beceri olarak ele alınabilir (Tatsuoka, 1995). Bu sayede herhangi bir bireyin, yerleştiği örtük sınıfa göre önceden tanımlanmış bazı becerilere sahip olup olmadığı ya da hangi olasılıkla sahip olabileceği ortaya konabilir. BTM uygulamaları Bilgisayarda Bireyselleştirilmiş Test (BBT) uygulamalarında da gerçekleştirilebilmektedir (Cheng, 2009). Literatürde BBT temelli BTM uygulamaları için çeşitli başlangıç, ilerleme ve sonlandırma kuralları önerilmiştir (Sorrel ve ark., 2020).

Son dönemde makine öğrenmesi yöntemlerinin geleneksel BBT uygulamalarına alternatif olabileceği düşünülmektedir. Özellikle, parametrik olmayan bir yaklaşım olması nedeniyle sıkça tercih edilen bir makine öğrenmesi algoritması olan karar ağaçlarının, ölçülen özelliğin boyut sayısının çok olduğu durumlarda efektif çözümler üretebildiğine

ilişkin çalışmalara son dönemde sıkça rastlanmaktadır (Gomez, Laria & Hernandez, 2019; Rodriguez ve ark. 2020; Zheng ve ark., 2020). Bu çalışmada ise BT-BBT uygulamasının makine öğrenmesi algoritmalarından karar ağacı algoritması kullanılarak gerçekleştirilebileceği yeni bir yöntem önerilmektedir. Söz konusu yöntemin uygulama basamaklarına aşağıda yer verilmiştir.

a. BBT uygulamalarının tamamında gerçekleştirilen gerçek uygulama öncesi havuz kalibrasyonu amacıyla yapılan ön uygulama yeni yöntem için de gerçekleştirilecektir. Bu defa amaç geleneksel uygulamanın aksine madde parametrelerini belirlemek değil, bireylerin örtük sınıflamalarını gerçekleştirecek karar ağacı modelini oluşturmaktır.

b. Modelin oluşturulması sürecinde havuzdaki maddeler aynı Q matris (beceri) yapısındaki maddeler birlikte olacak şekilde küçük madde öbeklerine ayrılmaktadır (3-4-5'li öbekler). Her bir öbek karar ağacında bir yaprak (node/düğüm) oluşturmaktadır.

c. Bir öğrenci için BT-BBT uygulaması algoritma tarafından karar verilmiş bir madde öbeğinin öğrenciye yöneltilmesi ile başlamaktadır. Öğrencinin bu öbekteki maddelerden alacağı toplam puana göre (kritik kesme noktaları algoritma tarafından belirlenmektedir) öğrenciye başka bir madde öbeği yöneltilmektedir. Bu işlem söz konusu öğrenci önceden oluşturulmuş karar ağacı modelinde bir örtük sınıfa yerleşinceye kadar devam ettirilir.

Yukarıda aşamaları verilen BT-BBT uygulamasını incelediğimizde uygulamaya ait başlangıç kuralı, ilerleme kuralı ve sonlandırma kurallarının tamamı Karar Ağacı algoritması tarafından belirlenmektedir. Uygulamada her öğrenciye yönlendirilen maddelerin farklı olması ve maddelerin değişken olması, bir sonraki yönlendirilecek sorunun öğrencinin önceki soruya verdiği cevaba göre interaktif olarak seçilmesi yönleriyle yeni önerilen yöntemin bir bireye uyarlanmış test özelliği taşıdığı söylenilebilir. Karar ağacına dayalı BT-BBT uygulamak için sırasıyla aşağıdaki işlem adımlarının izlenmesi gereklidir:

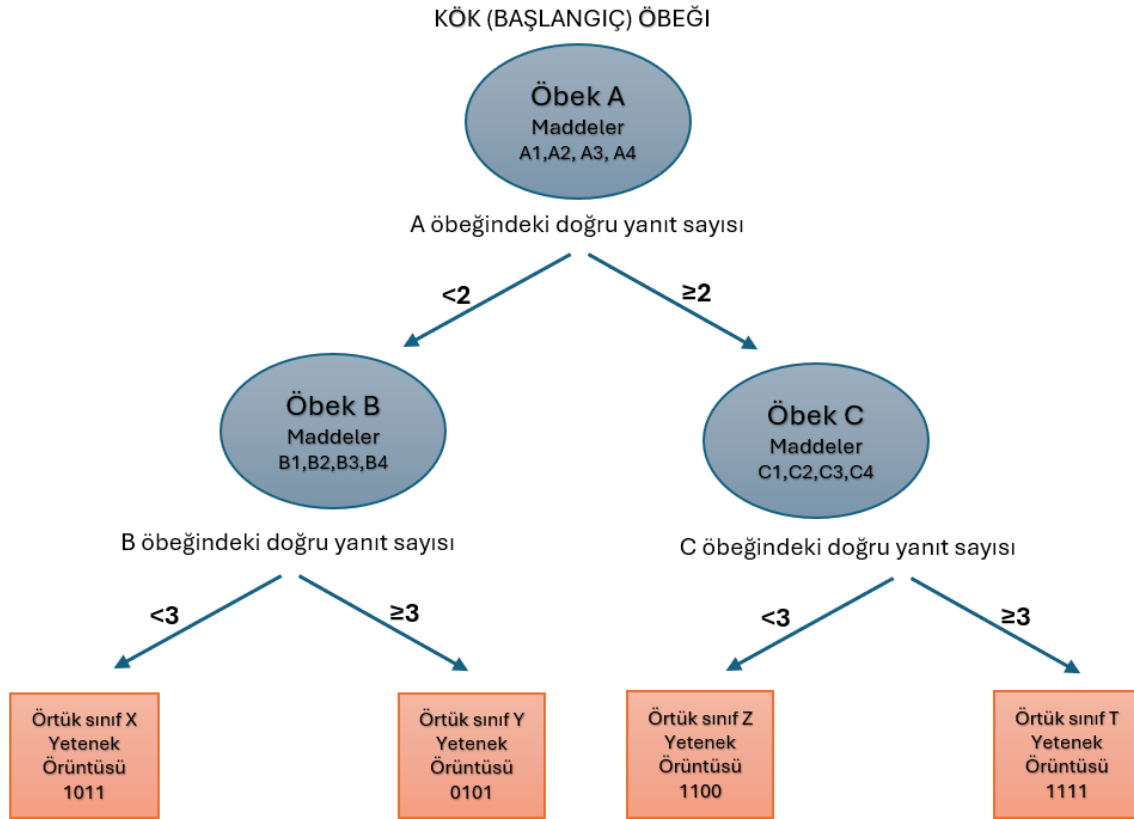
1. Madde havuzundaki maddelerin Q matrisi yapısına göre homojen gruplara ayrılması.

2. Aynı yapıya sahip maddelerden belirli büyüklükte madde topluluklarının oluşturulması (madde öbekleri).
3. Deneme uygulamasından elde edilen veride her bir öğrencinin her bir madde öbeğinden elde ettiği toplam puanların hesaplanıp kaydedilmesi.
4. Deneme verisi kullanılarak havuzdaki tüm maddeler analize dahil edilip öğrencilerin örtük sınıflarının kestirilmesi ve gerçek örtük sınıflar olarak kaydedilmesi.
5. Gerçek örtük sınıflar kategorik bağımlı değişken, öğrencilerin her bir öbekten aldığı toplam puanlar bağımsız değişken olacak şekilde bir karar ağacı modeli oluşturulması.

Yukarıdaki adımların sonunda elde edilen karar ağacı modeli, BT-BBT uygulamasının başlangıç, ilerleme ve sonlandırma kurallarının tamamını tanımlamaktadır. Katılımcıların tümü karar ağacının kök düğümündeki maddeleri cevaplandırarak teste başlayacaklardır. Katılımcılar, kök öbeğindeki maddelerden aldıkları puana (doğru yanıt sayısı) göre karar ağacı takip edilerek başka bir öbeğe yönlendirilecektir. Şekil 2’de derinliği 2 olan örnek bir BT-BBT karar ağacı modeli paylaşılmıştır.

Şekil 2

Karar Ağacına Dayalı Yöntemin Akış Şeması



Şekil 2'deki örnek diyagram incelendiğinde kök öbeğinden 2'den az doğru yanıt alan öğrenciler A öbeğindeki maddelerle, 2 ve daha fazla doğru yanıt alan öğrenciler B öbeğinde maddelerle teste devam edeceği görülmektedir. Karar ağacı modelinin her bir yaprağı (ucu) bir örtük sınıfa karşılık gelecektir. Bu sebeple, karar ağacı üzerinde ilerleyip bir yaprağa ulaşan öğrencinin ait olduğu örtük sınıf kestirilmiş olacaktır.

İlgili Araştırmalar

Kaplan ve ark. (2015) çalışmalarında BT-BBT uygulamalarında kullanılabilecek 2 farklı madde seçim yöntemi önermişlerdir. Önerilen bu madde seçim yöntemleri ile yaygın kullanılan Posterior Weighted Kullback Leibler (PWKL) yöntemini sınıflama tutarlılığı ve test uzunlukları açısından karşılaştırmışlardır. Elde edilen bulgularda iki yeni sınıflama

yönteminin benzer sonuçlar verdiği ve PWKL madde seçim yönteminden hem sınıflama tutarlılığı hem de test uzunlukları konusunda pozitif olarak ayrıştığı sonucuna ulaşmışlardır.

Madison ve Bradshaw (2015) çalışmalarında BT-BBT uygulamalarında kullanılan Q matrislerinin karmaşıklığının test güvenilirliği üzerine etkisini incelemişlerdir. Elde edilen bulgulara göre Q matris tanımlamalarında maddelerin sınırlı sayıda nitelik ile eşleştirildiği (basit yapılar) durumlarda test güvenilirliğinin daha yüksek olduğu sonucuna ulaşılmıştır.

Hsu ve Wang (2015) çalışmalarında değişken uzunluklu test sonlandırma kuralları ile sabit test uzunluğuna dayalı sonlandırma kurallarının performanslarını madde seçim kuralları perspektifinden inceleyerek karşılaştırmışlardır. Elde ettikleri bulgulara göre Kullback Leibler Information ve Posterior Weighted Kullback Leibler Information madde seçim yöntemlerini önermişlerdir. Ayrıca değişken test uzunluğuna dayalı (Minimum kesinlik) test sonlandırma kurallarının sabit test uzunluğuna dayalı sonlandırma kurallarına göre daha iyi çalıştığı sonucuna ulaşmışlardır.

Sorrel ve ark. (2020) çalışmalarında Double Kullback Leibler (DKL) ve G-DINA Global Discrimination Index (GDI) madde seçim yöntemlerini karşılaştırmışlardır. Elde edilen bulgulara göre DKL yönteminin sınıflama başarısını arttırdığı ama maruz kalma oranını da olumsuz etkilediği vurgulanmıştır. İki yöntem arasındaki sınıflandırma başarısı farkının test uzunluğu arttıkça kapandığı tespit edilmiştir. Test güvenliği göz önüne alındığında DKL yönteminin sonuçları düşük önem arz eden (low stakes) testlerde kullanılması önerilmiştir.

Delgado-Gomez, Laria ve Ruiz-Hernandez (2019), BBT oluşturmak için yeni bir yöntem olan Tree-CAT'i tanıtmaktadır. Bu yaklaşım, mevcut çalışmalardan farklı olarak, BBT ile karar ağaçları arasındaki benzerliklerden yararlanmaktadır. Tree-CAT, sınav katılımcılarının yetenek seviyelerini tahmin ederken ortalama kare hatayı en aza indirmeyi ve madde maruziyetini kontrol etmeyi sağlar. Bu süreç, karar ağaçları aracılığıyla, her dal ile ilişkili maddelerin seçilmesini sağlayan yenilikçi bir algoritma kullanarak doğrusal bir program çözümü ile aşamalı olarak gerçekleştirilir. Ayrıca, bu çalışma, anlık madde seçimi

veya testin sınırsız sayıda katılımcıya aynı anda uygulanması gibi maruziyet kontrolü gerektiren alternatif madde seçimi yöntemlerine kıyasla birçok avantaj sunmaktadır. Bu avantajlar, madde seçim yönteminin hesaplama açısından yoğun olduğu durumlarda bile doğru çevrimiçi BBT uygulamalarının mümkün olmasını sağlamaktadır.

Rodríguez-Cuadrado, Delgado-Gómez, Laria ve Rodríguez-Cuadrado (2020) çalışmasında, kişisel bilgisayarlarda yalnızca birkaç saniye içinde karar ağaçlarına dayanan BBT'ler oluşturmayı sağlayan Birleştirilmiş Tree-CAT yöntemini geliştirmişlerdir. Bu yöntem hem dağılım hem de gizli seviyenin tahminlerinin benzer olduğu dalları birleştirerek ağacın büyümesini kontrol etmektedir. Yapılan deneyler, bu önerilen yaklaşımın hesaplama süresini önemli ölçüde azalttığını ve en güncel tekniklerle karşılaştırıldığında benzer gizli seviye tahminleri sağladığını ortaya koymaktadır.

Bölüm 3

Yöntem

Bu bölümde araştırmanın yöntemine ilişkin açıklamalar yapılmıştır. Bu kapsamda araştırmanın türü, araştırmanın deseni, verilerin tanıtılması, verilerin türetilmesi, simülasyon koşulları ve analiz çıktıları ile ilgili bilgiler sunulmuştur.

Araştırmanın Türü

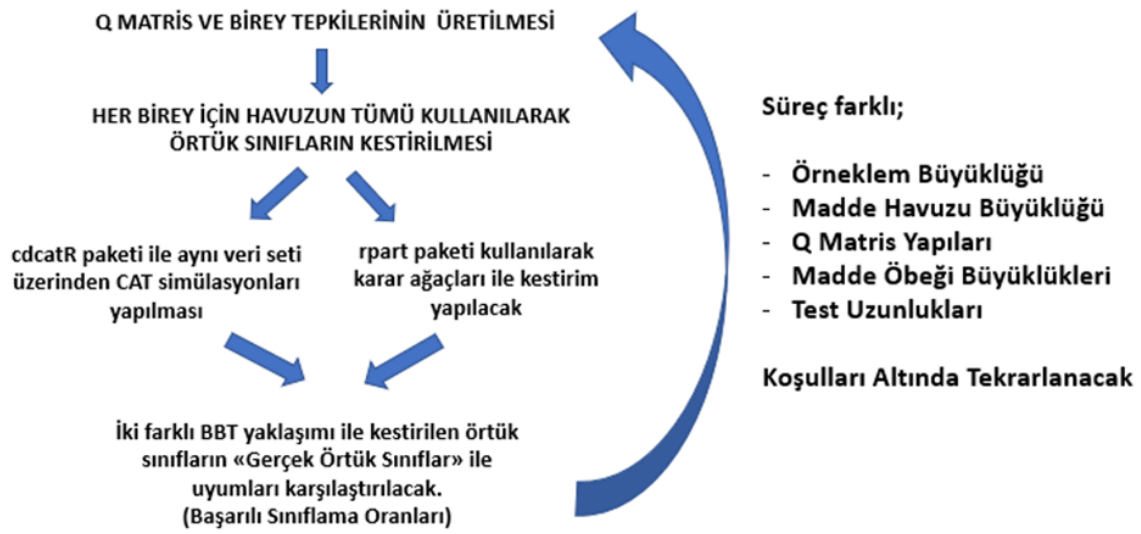
Araştırmada çeşitli BT-BBT yöntemlerini karşılaştırmak üzere yapay veriler kullanılarak çeşitli koşullar değiştirilerek analizler yürütülmüştür. Bu yönüyle araştırmanın bir Monte Carlo simülasyon çalışması olduğu söylenebilir (Fishman & Moore, 2013; Rubinstein & Kroese, 2016).

Araştırmanın Deseni

Bu bölümde araştırmada karar ağaçlarına dayalı BT-BBT uygulamalarının başarılı sınıflama oranları ile geleneksel BT-BBT yöntemlerine ait başarılı sınıflama oranlarının karşılaştırılmasına ilişkin oluşturulan araştırma deseni ile ilgili detaylı bilgiler sunulmuştur. Söz konusu desene ait uygulama adımları Şekil 3'te şematize edilerek açıklanmıştır.

Şekil 3

Araştırma Deseni Şeması



Araştırmada simülasyon çalışmalarının yürütülmesinde ilk adım simülasyon koşullarına uygun parametreler kullanarak Q matrisi ve 1-0 veri setinin türetilmesidir. Üretilen veri seti ve Q matrisi ile geleneksel yöntemler kullanarak bilişsel tanı analizi gerçekleştirilmekte ve her bir bireye ait örtük sınıf aidiyeti kestirilmektedir. Söz konusu örtük sınıfların kestirimi için yapılan analizlerde üretilen madde havuzunun tamamı kullanılmaktadır. Bu sebeple gerçekleştirilebilecek herhangi bir bireyselleştirilmiş test uygulamasından daha verimli sonuçlar vereceği söylenebilir. Bu sebeple belirlenen örtük sınıflar, araştırmanın devamında bireyselleştirilmiş uygulamalar ile farklı yöntemler kullanılarak elde edilecek sınıf aidiyetlerinin doğruluğunu değerlendirmek için referans olarak kullanılacaktır.

Üretilen veri setleri önce karar ağacı temelli yaklaşımla analiz edilmektedir. Bu kapsamda öncelikle maddelere ait 1-0 puanlar toplanarak simülasyon koşuluna uygun şekilde madde öbekleri elde edilmektedir. Madde öbekleri elde edilirken bir araya getirilen maddelerin aynı Q matris beceri yapısına sahip olması sağlanmaktadır. Bu süreç araştırmacı tarafından yazılan R kodları kullanılarak otomatik olarak gerçekleştirilmektedir. Madde öbekleri elde edildikten sonra her bireyin her madde öbeğinden elde ettiği toplam

puanlar hesaplanarak madde öbeği bazında yeni bir puan matrisi elde edilmektedir. Bu puan matrisinin son sütununa ise bireylere ait daha önceden kestirilmiş referans örtük sınıflar eklenmektedir. Söz konusu matristeki öbek puanlarını bağımsız değişken, örtük sınıflar bağımlı değişken olarak kullanılarak örtük sınıfları en iyi şekilde tahmin eden karar ağacı modeli “rpart” paketinin (Therneau & Atkinson, 2019) aynı isimli fonksiyonu kullanılarak elde edilmektedir. Karar ağacı oluşturulurken bireylere yönlendirilecek madde sayıları öbek büyüklükleri ve test uzunluğu koşulları göz önünde bulundurularak yazılan kodlar yardımı ile otomatik olarak kontrol edilmektedir. Karar ağacı oluşturulurken kullanılan diğer özellik ve parametreler ile ilgili bilgi verilerin analizi bölümünde sunulmuştur.

Karar ağacı modeli oluşturulduktan sonra modelin örtük sınıf aidiyetlerini tahminleme başarısı kontrol edilip beklenen sınıflar ve beklenen sınıflar arasındaki eşleşme oranı elde edilip raporlanmaktadır. Aynı zamanda örtük sınıf tahminleri dışında bireylerin sahip oldukları becerilerin ayrı ayrı tahmini de bu ağaç modeline göre gerçekleştirilmekte ve referans beceri durumları ile karşılaştırılarak eşleşme oranları hesaplanmaktadır.

Karar ağacı modeli ile oluşturulan bireyselleştirilmiş test simülasyonunun başarısı geleneksel yöntemler ile gerçekleştirilmiş BT-BBT simülasyonlarının başarısı ile karşılaştırılmak istenmektedir. Bu sebeple “cdcatR” paketi (Sorrel ve ark., 2021) yardımı ile aynı veri setleri için alan yazına dayalı farklı ilerleme kuralları kullanılarak aynı test uzunluğu koşullarında BT-BBT simülasyonları yürütülmüştür.

Söz konusu yöntemlerin karşılaştırılmasında bireylere ait beklenen örtük sınıflar ve beceri aidiyetleri olarak referans alınacak değerler, madde havuzundaki tüm maddeler kullanılarak geleneksel yöntemler ile yapılan kestirimlerdir. Dolayısı ile karar ağacına dayalı önerilen yöntemin geleneksel bireyselleştirilmiş test uygulamalarına göre verimliliğini ortaya koymak adına başarılı sınıf ve beceri tahmin oranları karşılaştırılmıştır.

Gerçek Verilere Dayalı Yürütülen Simülasyonlar

Araştırmanın ikinci alt probleminde yapılan simülasyonlarda yapay olarak üretilen öğrenci tepkileri, madde parametreleri ve Q matrisleri kullanılmamıştır. Bunun yerine bireyselleştirilmiş test simülasyonlarında kullanılacak s ve g parametreleri gerçek bir öğrenci yanıt matrisinden faydalanılarak GDINA modeli altında kestirilmiştir. Söz konusu gerçek verinin elde edildiği maddelere ait gerçek Q matrisleri uzman görüşü alınarak 3 beceri içerecek şekilde hazırlanmıştır. Daha sonra “cdcatR” R paketi yardımıyla yürütülen bireyselleştirilmiş test simülasyonlarında fonksiyona yapay olarak üretilmiş madde parametreleri ve Q matrisleri yerine söz konusu gerçek veriye dayalı madde parametreleri ve Q matrisi argüman olarak dahil edilmiştir.

Verilerin Tanıtılması

Bu çalışmada karşılaştırılan BT-BBT yöntemlerinin çeşitli koşullar altında performanslarını gözlemlemek amacıyla 144 farklı koşulda simülasyon verileri üretilmiştir. Bunun yanı sıra yapay veriler ile elde edilen sonuçları desteklemek amacı ile gerçek bir veri seti de kullanılmıştır. Bu bölümde kullanılan gerçek veri seti tanıtılmış ayrıca üretilen veri setlerinin üretim koşulları hakkında bilgiler sunulmuştur.

Verilerin Üretilmesi

Çalışmada 144 farklı koşulda 20’şer replikasyon gerçekleştirilmiştir. Buna göre toplamda 2880 farklı iki kategorili veri seti üretilmiştir. Ayrıca üretilen her bir veri seti için 3 beceri tanımlanmış birer Q matris üretilmiştir. Q matrislerin üretimi “simcdm” paketinin (Balamuta ve ark., 2019) “sim_q_matris” fonksiyonu kullanılarak gerçekleştirilmiştir. Elde edilen Q matrisleri kullanılarak “CDM” paketinin (Robitzsch ve ark., 2020) “sim.gdina” fonksiyonu aracılığıyla iki kategorili veri matrisleri elde edilmiştir. Veri setlerinin üretiminde kullanılan s ve g parametre değerleri araştırmada değişimlenen koşullardan biri olduğundan gerekli bilgiler simülasyon koşulları başlığı altında sunulmuştur.

Her bir veri seti 1000 hipotetik öğrenciden oluşmaktadır. 1000 kişilik simülasyon koşulu alan yazın dikkate alınarak tercih edilmiştir (Hsu & Wang, 2015).

Gerçek Veri Setinin Tanıtılması

Araştırmada simülasyon aşamasında elde edilen sonuçların desteklenmesi amacı ile gerçek bir veri seti ile BT-BBT uygulamaları gerçekleştirilmiştir. Bu kapsamda 3243 altıncı sınıf öğrencisinin katıldığı 20 maddelik bir matematik testine ait iki kategorili puanlanmış veri seti kullanılmıştır. Kullanılan verinin Cronbach Alfa iç tutarlılık katsayısı 0.897 olarak kestirilmiştir. Söz konusu 20 maddeye ait Q matrisi araştırmacı dahil 3 alan uzmanı tarafından görüş birliği ile oluşturulmuştur. Oluşturulan Q matrisi Tablo 1’de sunulmuştur.

Kullanılan gerçek veri setinden elde edilen s parametrelerinin minimum değeri 0,121 iken maksimum değeri 0,332’tür. Kestirilen g parametrelerinde ise minimum değerin 0,065 ve maksimum değerin 0,237 olduğu gözlenmiştir. Tablo 2’de verilen ve literatürde yüksek ve düşük ayıricılık düzeyine sahip maddeler için eşik değer olarak önerilen değerler incelendiğinde, gerçek verilerden elde edilen madde parametrelerinin genel olarak ayıricılık düzeylerinin yüksek olduğu söylenilebilir.

Tablo 1

Q Matris

Madde Numarası		Sözlü	Tek ya da çok adımlı
		matematiksel	matematiksel
	Kesirler ile işlem	ifadeleri	işlemleri
	yapabilme	anlamlandırma	gerçekleştirme
1	0	1	1
2	0	0	1
3	0	0	1
4	0	1	0

5	0	1	1
6	0	0	1
7	0	1	1
8	0	1	0
9	0	0	1
10	1	0	0
11	0	1	0
12	0	1	1
13	0	1	0
14	0	1	0
15	0	1	0
16	1	0	0
17	1	0	0
18	1	0	0
19	1	1	0
20	1	1	0

Q matriste kullanılan 3 beceri TIMSS 2019 Matematik alanı yeterlik düzeylerine ait beceri tanımları incelenerek tercih edilmiştir. Buna göre 20 maddenin 6'sı kesirler ile işlem yapabilme becerisini, 12'si sözlü matematiksel ifadeleri anlamlandırma becerisini, 8 tanesi ise tek ya da çok adımlı matematiksel işlemleri yapabilme becerisini gerektirmektedir.

Simülasyon Koşulları

Araştırmada yürütülen simülasyon koşulları kapsamında geliştirilen veri üretimi ve analiz süreçlerinde bazı koşullar sabit olarak kullanılırken, bazı koşullar ise değişimlenmiştir. Sabit bırakılan koşullar ile ilgili verilerin üretilmesi ve verilerin analizi bölümlerinde detaylı açıklamalar yapılmıştır. Çalışmada değişimlenen koşullar ise aşağıda listelenmiştir.

- Karar ağacı yönteminde kullanılan madde öbeklerinin büyüklükleri
- Test uzunlukları
- Bilişsel tanı modellerine dayalı üretilen iki kategorili maddelerin ayırıcılık güçleri (s ve g parametreleri)
- Madde havuzu büyüklüğü
- BT-BBT için kullanılan ilerleme kuralı

Karar Ağacı Yönteminde Kullanılan Madde Öbek Büyüklükleri

Araştırmada kullanılan karar ağacı algoritmasına dayalı yöntemde karar ağacına ait düğümler (node) tek birer maddeyi temsil etmek yerine madde öbekleri kullanılarak oluşturulmuştur. Bu durumun bir sebebi her bir düğümden ayrı ayrı verilen büyüktür-küçüktür kararı için gerekli puan açıklığının yaratılmasıdır. Bir diğer sebep ise karar ağacının karmaşıklığının (complexity) azaltılmasıdır. Söz konusu öbeklerin oluşturulma süreci ile ilgili bilgi verilerin analizi bölümünde sunulmuştur.

Madde öbek büyüklükleri alan yazında yaygın kullanılan test uzunluğu koşulları da göz önünde bulundurularak 2, 3 ve 4'erli madde öbekleri olmak üzere 3 farklı koşulda değişimlenmiştir.

Test Uzunlukları

Araştırmada kullanılan bireyselleştirilmiş test yöntemlerinin başarısına ilişkin kullanılan değerlendirme ölçütleri başarılı örtük sınıf tahmini oranı ve başarılı beceri tahmini oranıdır. Bu değerleri etkileyen en önemli değişkenlerden biri de bireyselleştirilmiş test uygulaması boyunca her bir bireye uygulanan madde sayısı yani test uzunluklarıdır. Bireylere uygulanan test uzunlukları arttıkça yapılan sınıflamaların başarı oranı artmaktadır. Bu sebeple alan yazında bireyselleştirilmiş test uzunluklarının değişimlendiği çalışmalara rastlanmaktadır.

Bu çalışmada karar ağacı yöntemi çerçevesinde maddeler bireylere madde öbekleri halinde yönlendirilmiştir. Farklı koşullarda madde öbek büyüklükleri 2, 3 ve 4 olarak seçildiğinden test uzunluklarının söz konusu öbek büyüklüklerine kolay bölünebilecek sayılardan tercih edilmiştir. Ayrıca BT-BBT simülasyon çalışmalarında 10 ve 20 maddelik test uzunluklarının tercih edilen koşullardan olduğu gözlemlendiğinden (Chen ve ark., 2012; Wang, 2013; Hsu ve ark., 2013; Kaplan ve ark., 2015) bu uzunluklara yakın değerler belirlenmiştir. Bu kapsamda araştırmada test uzunlukları 12 madde, 18 madde ve 24 madde olmak üzere 3 farklı değer altında değişimlenmiştir.

Bilişsel Tanı Modellerine Dayalı Üretilen İki Kategorili Maddelerin Ayıricılık Güçleri

BT-BBT uygulamalarında sınıflama başarısı üzerinde etkili olduğu için ilgili simülasyon çalışmalarında değişimlenen en önemli etkenlerden biri de s ve g parametreleridir (Chen ve ark., 2012; Wang, 2013). Bu kapsamda havuzda kullanılan maddelere ait s ve g (slipping ve guessing) parametreleri tek biçimli (uniform) dağılımdan belirli koşullar altında üretilmiştir. Söz konusu koşullar ilgili alan yazında benzer simülasyon çalışmalarında kullanılan şartlar referans alınarak tercih edilmiştir (Cheng, 2010; Chen ve ark., 2012; Kaplan ve ark., 2015; Wang ve ark., 2020). Parametrelerin üretim şartları Tablo 2’de sunulmuştur.

Tablo 2

Madde Parametrelerinin Üretim Şartları

	s parametresi		g parametresi	
	Minimum	Maksimum	Minimum	Maksimum
Düşük Ayıricılık	0,15	0,4	0,1	0,3
Yüksek Ayıricılık	0,1	0,3	0,05	0,2

Tablo 2 incelendiğinde çalışmada düşük ve yüksek ayıricılık gücüne sahip olmayı temsil eden iki farklı madde parametresi üretim koşulu seti kullanıldığı görülmektedir. Düşük ayıricılık gücüne sahip maddelerin üretimi için s parametreleri tek biçimli dağılımdan minimum 0,15 maksimum 0,4 değerleri arasından, g parametreleri ise minimum 0,1

maksimum 0,3 değerleri arasından rastgele seçilmiştir. Yüksek ayırıcılık gücüne sahip maddeleri simüle etmek için ise s parametreleri minimum 0,1 maksimum 0,3 aralığından seçilirken, g parametreleri ise minimum 0,05 maksimum 0,2 değer aralığından elde edilmiştir.

Madde Havuzu Büyüklüğü

BT-BBT ile ilgili yapılan simülasyon çalışmalarında sıkça değişimlenen ve sınıflama başarısı üzerinde etkili olan bir diğer değişken ise madde havuzu büyüklüğüdür. Alan yazın incelendiğinde 300 maddelik havuz büyüklüğünün sıklıkla kullanıldığı görülmüştür (Cheng, 2010; Guo & Zeng, 2019; Hsu ve ark., 2013; Hsu & Wang, 2015; Liu ve ark., 2013). Ayrıca Chen ve Xin (2012) çalışmalarında 120 maddelik bir madde havuzu kullanmışlardır. Bu çalışmada ise literatürde yaygın olarak görülen 300 maddelik havuzlara göre daha küçük olan 100 ve 200 maddelik iki havuz koşulu tercih edilmiştir.

Araştırmada madde maruz kalma sıklığı ya da içerik dengeleme gibi madde havuzu kısıtlayıcı teknikler kullanılmadığından geniş madde havuzu koşulları tercih edilmemiştir. Ayrıca yöntem karşılaştırması yapılan simülasyon çalışmalarında yöntemlerin yüksek başarıya ulaşabileceği koşulların tercih edilmesinin yöntemler arasındaki olası farklılıkların tespit edilmesini güçleştirebileceği düşünülmektedir. Son olarak değişimlenen koşullar arasında madde kalitesini (ayırıcılık gücünü) temsil eden s ve g parametrelerinin bulunması, söz konusu parametrelerin tek biçimli dağılımdan üretildiği ve test uzunluklarının 12, 18 ve 24 madde olarak değişimlendiği göz önüne alındığında çok geniş madde havuzlarının madde ayırıcılık gücü değişimlerinin sonuçlarının gözlenmesini güçleştireceği düşünülmüştür.

BT-BBT Madde Seçim Kuralı

Araştırmada karar ağacı yönteminin sınıflandırma başarısını karşılaştırmak adına geleneksel BT-BBT uygulamaları da simüle edilmiştir. Bu kapsamda karar ağacı temelli yöntemin yanında literatürde yaygın kullanılan iki farklı ilerleme kuralı ve bunlara ek olarak

havuzdan madde seçimini rastgele gerçekleştiren bir koşul kullanılmıştır. Bu kapsamda karar ağacı, G-DINA Global Discrimination index (GDI) (Kaplan ve ark., 2015), Posterior Weighted Kullback Leibler (PWKL) (Cheng, 2009) ve rastgele madde seçim yöntemleri olmak üzere 4 farklı ilerleme yöntemi simülasyon koşullarına eklenmiştir.

Yukarıda açıklandığı gibi madde öbek büyüklükleri (3 koşul), test uzunlukları (3 koşul), madde ayırıcılık güçleri (2 koşul), madde havuzu büyüklüğü (2 koşul) ve madde seçim kuralı (4 koşul) değişimlenmiştir. Tablo 3'te değişimlenen tüm değişkenler ve koşulları sunulmuştur.

Tablo 3

Değişimlenen Simülasyon Koşulları

Değişimlenen Değişken	Koşul 1	Koşul 2	Koşul 3	Koşul 4
Öbek Büyüklüğü	2	3	4	--
Test Uzunluğu	12	18	24	--
Madde Ayırıcılık Düzeyi	Düşük	Yüksek	--	--
Madde Havuzu	100	200		
İlerleme Kuralı	Karar Ağacı	GDI	PWKL	Rastgele

Tablo 3'te görüldüğü gibi araştırmada $3 \times 3 \times 2 \times 2 \times 4 = 144$ farklı koşulda analizler gerçekleştirilmiştir. Söz konusu 144 koşulun her biri için 20 replikasyon uygulanmıştır. Böylece bu çalışmada toplamda $144 \times 20 = 2880$ adet BT-BBT simülasyonu gerçekleştirilmiştir.

Verilerin Analizi

Araştırmada kullanılan verilerin hazırlanması, türetilmesi ve analizler R programlama dilinde yazılan kodlar ve bazı paket programlar kullanılarak gerçekleştirilmiştir. Karar ağacı modelleri ve oluşturulan modeller kullanılarak yapılan örtük sınıf kestirimleri R dilinin 'rpart' paketi (Therneau & Atkinson, 2019) kullanılarak yapılmıştır.

Bilişsel tanı modeline dayalı bireyselleştirilmiş test simülasyonları için cdcR (Sorrel & Abad, 2021) paketinden faydalanılmıştır. Simülasyonlarda kullanılan Q matrisleri “simcdm” (Balamuta ve ark., 2019) paketinden faydalanılarak türetilmiştir. Araştırmada üretilen verilerin %25’i eğitim %75’i test veri seti olarak kullanılmıştır. Böylece aşırı öğrenme problemi olup olmadığı kontrol edilmiş ve aşırı öğrenme problemi olmadığı tespit edilmiştir.

Araştırmada analizler sonucunda kullanılan yöntemlerin kestirim başarılarını ortaya koymak için iki farklı gösterge kullanılmıştır. Bu göstergeler, Başarılı Örtük Sınıf Tahmini Oranı (BSTO) ve Başarılı Beceri Tahmini Oranı’dır (BBTO). Aşağıda bu göstergelerin elde edilen yöntemlerine ilişkin detaylı açıklamalar verilmiştir.

Başarılı Örtük Sınıf Tahmini Oranı (BSTO)

BT-BBT uygulamalarının temel amacı bireyin yanıt örüntülerini kullanarak, bireyi 2K (K: Q matriste kullanılan beceri sayısı olmak üzere) sayıda örtük sınıftan birine atamaktır. Simülasyon süreci öncesi birey tepkileri üretirken bireylerin ait olduğu gerçek örtük sınıflar bilinmektedir. Bu sayede gerçek örtük sınıflar ile simülasyon sonucu yerleştirilen örtük sınıfın uyumlu olup olmadığı kontrol edilebilmektedir. Başarılı örtük sınıf tahmin oranı basitçe, örtük sınıf sınıflaması doğru yapılan birey sayısının tüm gruptaki birey sayısına oranı olarak düşünülebilir. Bu başarının sağlanması için bireyin, Q matristeki K sayıda beceriye sahip olma durumlarının tamamını doğru kestirmiş olması gerekmektedir. BT-BBT uygulamalarının temel işlevini yerine getirip getirmediğini ortaya koyması açısından araştırmanın bu çıktısının son derece önemli olduğu düşünülmektedir.

Başarılı Beceri Tahmini Oranı (BBTO)

Araştırmanın bir diğer çıktısı olan başarılı beceri sınıflaması oranı bireyin Q matrisindeki K tane beceriden kaçında gerçek ve kestirilen sınıflamanın uyumlu olduğuna ait bir göstergedir. N sayıda birey ve K sayıda beceri türü olan bir BT-BBT uygulamasında toplam $N \times K$ tane beceri sınıflaması yapılacaktır. Bu sınıflamalardan doğru olanların

sayısının, tüm sınıflamalara ($N \times K$ adet) oranı başarılı beceri sınıflaması oranını temsil etmektedir.

Örneğin i. bireyin gerçek beceri aidietleri $K_1=1$, $K_2=0$, $K_3=0$, $K_4=0$, $K_5=1$ olsun. Aynı bireyin BT-BBT uygulaması sonucunda kestirilen beceri aidietleri ise $K_1=1$, $K_2=0$, $K_3=0$, $K_4=0$, $K_5=0$ olsun. Birey hakkında yapılan beş beceri kestiriminden dördü (K_1 , K_2 , K_3 , K_4 ,) doğru, biri (K_5) hatalı olduğu için bireyin başarılı beceri sınıflaması oranı $4/5 = 0,80$ olarak hesaplanacaktır. Tüm bireylere ait başarılı beceri sınıflaması oranlarının ortalama değeri ise BT-BBT uygulamasının başarılı beceri sınıflandırması olarak kabul edilecektir.



Bölüm 4

Bulgular, Yorumlar ve Tartışma

Bu bölümde araştırmanın bulgularına ilişkin bilgiler araştırma problemleri akışına dayalı olarak sunulacaktır.

1. Alt Probleme İlişkin Bulgular

Bu çalışmanın birinci alt probleminde simülasyon verisinde karar ağacına dayalı uygulama ve BT-BBT uygulamaları ile elde edilen başarılı beceri ve sınıf tahmini oranları araştırılmıştır. Çalışmada karar ağacı ve BT-BBT uygulamaları 144 farklı simülasyon koşulunda 20'şer replikasyon altında karşılaştırılmaktadır. Birinci araştırma probleminde değişimlenen koşullardan bağımsız olarak iki farklı yaklaşımın tüm koşullardan elde edilen başarılı tahmin oranlarının karşılaştırılması hedeflenmiştir. Bu hedef doğrultusunda Tablo 4'te simülasyon analizleri sonucu elde edilen başarılı örtük sınıf tahmini oranları sunulmuştur.

Tablo 4*Farklı Simülasyon Koşullarına Göre Başarılı Örtük Sınıf Tahmini Oranları*

Ayrılcılık	Düzeyi	Test	uzunluğu	Havuz	Büyüklüğü	Karar Ağacı (Öbek Büyüklüğü)			GDI	PWKL	Rastgele
						2	3	4			
Düşük	12		100	0,812	0,784	0,676	0,851	0,767	0,578		
			200	0,853	0,802	0,706	0,890	0,793	0,563		
		18	100	0,964	0,905	0,879	0,862	0,769	0,571		
			200	0,967	0,921	0,895	0,883	0,795	0,563		
		24	100	0,997	0,974	0,940	0,873	0,784	0,574		
			200	0,998	0,981	0,948	0,885	0,801	0,571		
	12		100	0,945	0,887	0,790	0,959	0,890	0,711		
			200	0,951	0,889	0,841	0,963	0,899	0,710		
		18	100	0,995	0,975	0,967	0,953	0,887	0,706		
			200	0,996	0,984	0,969	0,962	0,894	0,709		
		24	100	0,998	0,996	0,992	0,952	0,886	0,708		
			200	1,000	0,998	0,996	0,965	0,892	0,701		
Ortalama						0,956	0,925	0,883	0,917	0,838	0,639

Tablo 4'ün en alt satırında karar ağacına dayalı 2, 3 ve 4 maddelik öbekler kullanılarak yapılan uygulamalara ve GDI, PWKL ve Rastgele madde seçim yöntemleri kullanılan BT-BBT uygulamaları ile tüm koşullarda elde edilen başarılı örtük sınıf tahmini oranlarının ortalama değerleri sunulmuştur. Buna göre 2 maddelik öbekler kullanılan karar

ağacı uygulaması 12 koşulda ortalama 0,956'lık bir başarılı örtük sınıf tahmini oranı elde etmiştir. 3 maddelik öbeklerin kullanıldığı uygulamalarda ortalama 0,925'lik başarılı sınıflama oranı gözlenirken, 4 maddelik öbeklerde ise bu oran 0,883 olarak elde edilmiştir. Geleneksel BT-BBT uygulamalarına ilişkin ise 3 farklı madde seçim yöntemi ile analizler gerçekleştirilmiştir. Buna göre GDI yöntemi ile 12 farklı koşulda elde edilen başarılı örtük sınıf tahmini ortalaması 0,917 olarak elde edilmiştir. Bu oran PWKL madde seçim yöntemi ile 0,838 iken rastgele madde seçim yönteminde ise 0,639 olarak belirlenmiştir. Elde edilen bu altı değer incelendiğinde 12 farklı simülasyon koşulunun ortalamalarında en yüksek başarılı örtük sınıflama oranının 2 maddelik öbeklerin kullanıldığı karar ağacı temelli uygulamadan elde edildiği görülmüştür. İkinci en yüksek oran 3 maddelik öbeklerin kullanıldığı uygulamalar olurken üçüncü en yüksek oranı ise GDI ilerleme yönteminin kullanıldığı BT-BBT uygulamalarından elde edilmiştir. Tüm yöntemler arasında rastgele madde seçimi ile gerçekleştirilen BT-BBT uygulamaları ise beklenildiği gibi en düşük başarılı sınıf tahmini oranlarını elde etmiştir. Tablo 5'te ise tüm simülasyon koşullarında 6 farklı uygulama ile elde edilen ortalama başarılı beceri tahmini oranları sunulmuştur.

Tablo 5*Farklı Simülasyon Koşullarına Göre Başarılı Beceri Tahmini Oranları*

Ayıricılık	Test	Havuz	Karar Ağacı			GDI	PWKL	Rastgele
	Uzunluğu	Büüklüğü	2	3	4			
Düşük	12	100	0,921	0,911	0,862	0,941	0,910	0,818
		200	0,938	0,919	0,872	0,957	0,922	0,804
	18	100	0,985	0,961	0,950	0,946	0,911	0,811
		200	0,986	0,966	0,956	0,954	0,923	0,806
	24	100	0,999	0,989	0,975	0,950	0,918	0,810
		200	0,999	0,992	0,978	0,955	0,925	0,810
	12	100	0,976	0,956	0,916	0,984	0,961	0,882
		200	0,979	0,958	0,936	0,986	0,964	0,881
	18	100	0,998	0,990	0,987	0,982	0,959	0,878
		200	0,999	0,994	0,988	0,986	0,962	0,880
	24	100	0,999	0,999	0,997	0,982	0,959	0,878
		200	1,000	0,999	0,998	0,987	0,961	0,875
Ortalama			0,982	0,970	0,951	0,968	0,940	0,844

Tablo 5 incelendiğinde rastgele madde seçimine dayalı BT-BBT koşulları dışında tüm simülasyon koşullarında oldukça yüksek başarılı beceri sınıflaması oranları elde edildiği görülmektedir. 4 maddelik öbeklere dayalı ve 12 maddelik test uzunluklarına dayalı karar ağacı uygulamaları dışında tüm koşullarda 0,90 üzerinde başarılı beceri sınıflaması oranları

elde edilmiştir. Bunun yanı sıra 12 maddelik test uzunluğuna dayalı 4 farklı koşul dışında tüm koşullarda en yüksek başarılı beceri sınıflaması oranları 2 maddelik öbekler kullanılarak yürütülen karar ağacı uygulamalarından 0,982 olarak elde edilmiştir. 3 maddelik öbeklerin kullanıldığı karar ağacı uygulaması ve GDI ilerleme kuralının kullanıldığı BT-BBT uygulamasından ise oldukça yakın başarılı beceri tahmini oranları elde edilmiştir.

1.1. Alt Problemine İlişkin Bulgular

Bu alt problemde karşılaştırılan yöntemlerin yapay olarak üretilen maddelerin farklı ayırıcılık düzeyi koşullarındaki performanslarını incelemek amaçlanmıştır. Tablo 6'da düşük ve yüksek ayırıcılık düzeylerinde üretilen maddeler ile gerçekleştirilen karar ağacı ve BT-BBT uygulamalarının başarılı örtük sınıflama oranları raporlaştırılmıştır. Söz konusu oranlar hesaplanırken Tablo 4'te yer alan yüksek ve düşük ayırıcılık düzeyindeki maddelerin kullanıldığı koşullardan elde edilen BSTO değerlerinin ayrı ayrı ortalamaları alınmıştır.

Tablo 6

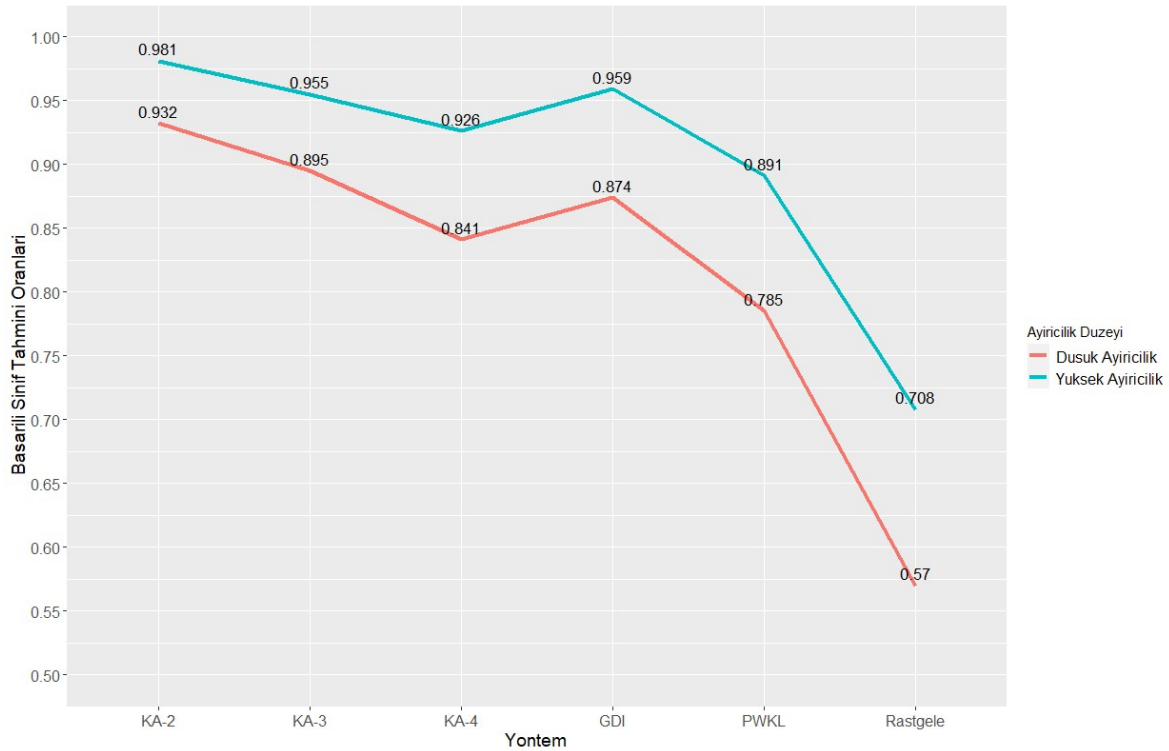
Farklı Madde Ayırıcılık Düzeylerine Göre Başarılı Sınıf Tahmini Oranları

	Karar Ağacı (Öbek Büyüklüğü)			GDI	PWKL	Rastgele
	2	3	4			
Düşük Ayırıcılık	0,932	0,895	0,841	0,874	0,785	0,570
Yüksek Ayırıcılık	0,981	0,955	0,926	0,959	0,891	0,708

Tablo 6 incelendiğinde karar ağacı ve BT-BBT temelli altı uygulamanın tümünde yüksek ayırıcılığa sahip maddeler kullanılarak yapılan uygulamalarda önemli derecede daha yüksek başarılı sınıf tahmini oranları elde edildiği görülmektedir. En yüksek başarılı sınıflama oranlarının 2 maddelik öbeklerin kullanıldığı karar ağacı uygulamalarında elde edildiği tespit edilmiştir.

Şekil 4

Ayırıcılık Düzeylerine Göre Başarılı Sınıf Tahmini Oranları Grafiği



Şekil 4 incelendiğinde, tüm madde seçimi koşulları için yüksek ayırıcılık gücüne sahip maddelerden oluşan madde havuzlarının kullanıldığı koşullarda başarılı sınıflama oranlarının göreceli olarak daha yüksek olduğu gözlenmiştir. Başarılı sınıflama oranlarının yüksek ve düşük ayırıcılık koşulları arasındaki gösterdiği farklılık madde seçimi yöntemlerine göre değişkenlik göstermektedir. Bu farkın 0,138 ile 0,049 arasında değiştiği gözlenmiştir. Şekil 4'te KA-3 ve GDI yöntemlerinin hem yüksek hem de düşük ayırıcılık düzeyi koşullarında benzer BSTO değerleri elde ettiği gözlenmiştir. Araştırmada kullanılan altı yöntemin tümünde hem yüksek hem de düşük ayırıcılık düzeyine sahip madde koşullarında rastgele madde seçim yönteminin en düşük BSTO değerlerini elde ettiği saptanmıştır.

Tablo 7'de düşük ve yüksek madde ayırıcılık gücü koşulları altında elde edilen başarılı beceri tahmini oranları (BBTO) sunulmuştur. Söz konusu değerler Tablo 5'teki

düşük ve yüksek ayırıcılık düzeyindeki maddelerin kullanıldığı koşullardan elde edilen BBTO değerlerinin ayrı ayrı ortalamaları alınarak hesaplanmıştır.

Tablo 7

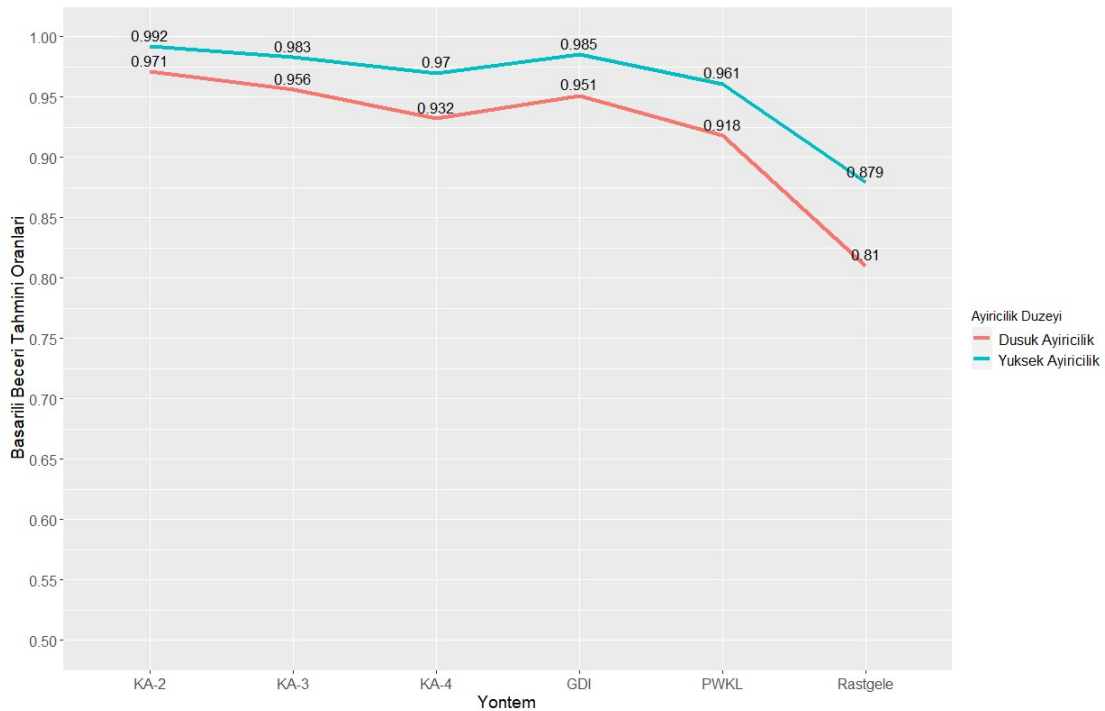
Farklı Madde Ayırıcılık Düzeylerine Göre Başarılı Beceri Tahmini Oranları

Ayırıcılık Düzeyi	Karar Ağacı			GDI	PWKL	Rastgele
	2	3	4			
Düşük	0,971	0,956	0,932	0,951	0,918	0,810
Yüksek	0,992	0,983	0,970	0,985	0,961	0,879

Tablo 7 incelendiğinde altı koşulun tümünde madde ayırıcılık düzeylerinin artışının başarılı beceri tahmini oranlarında artışlara sebep olduğu görülmektedir. Bu artışın başarılı beceri sınıflama oranlarının düşük olduğu koşullarda belirgin olduğu gözlenmiştir. Şekil 5'te maddelerin ayırıcılık düzeylerine göre Başarılı Beceri Tahmini Oranları sunulmuştur.

Şekil 5

Ayırıcılık Düzeylerine Göre Başarılı Beceri Tahmini Oranları Grafiği



Şekil 5 açıkça göstermektedir ki tüm madde seçim yöntemleri altında yüksek ayırıcılık düzeyine sahip maddelerin kullanıldığı koşullardan elde edilen BBTO değerleri düşük ayırıcılık düzeyine sahip koşullara göre daha yüksektir. Hem yüksek hem de düşük ayırıcılık düzeyine sahip maddelerin kullanıldığı koşullarda en yüksek BBTO değerlerini ikili madde öbeklerinin kullanıldığı karar ağacına dayalı yöntem (KA-2) elde etmiştir. Ardından KA-3 ve GDI yöntemlerinin BBTO oranlarının benzer düzeyde olduğu gözlenmiştir. Araştırmada kullanılan altı yöntemin tümünde hem yüksek hem de düşük ayırıcılık düzeyine sahip madde koşullarında rastgele madde seçim yönteminin en düşük BBTO değerlerini elde ettiği saptanmıştır.

1.2. Alt Probleme İlişkin Bulgular

Bu alt problemde karşılaştırılan yöntemlerin farklı sabit test uzunluğu koşullarında elde ettiği başarılı örtük sınıf tahmini oranları karşılaştırılmıştır. Sabit 12, 18 ve 24 maddelik test uzunluklarında 4'er farklı koşulda elde edilen oranların ortalamaları Tablo 8'de sunulmuştur.

Tablo 8

Farklı Test Uzunluklarına Göre Başarılı Sınıf Tahmini Oranları

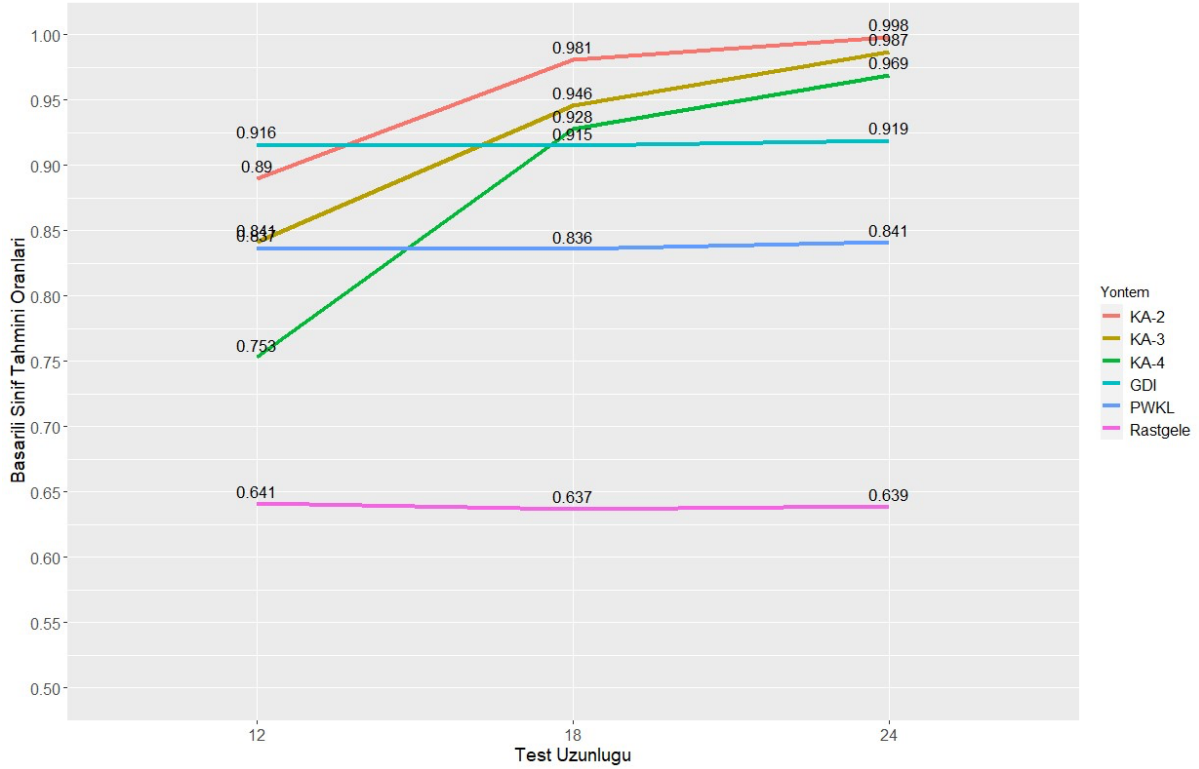
Test Uzunluğu	Karar Ağacı (Öbek Büyüklüğü)			GDI	PWKL	Rastgele
	2	3	4			
12	0,890	0,841	0,753	0,916	0,837	0,641
18	0,981	0,946	0,928	0,915	0,836	0,637
24	0,998	0,987	0,969	0,919	0,841	0,639

Tablo 8 incelendiğinde karar ağacı temelli uygulamaların BT-BBT uygulamalarına kıyasla test uzunluklarından daha fazla ve pozitif yönde etkilendiği görülmektedir. 12 maddelik test uzunluğu koşullarında GDI ilerleme kuralına dayalı BT-BBT uygulaması en

yüksek oranları yakalamıştır. Şekil 6'da farklı test uzunluğu koşullarında madde seçimi yöntemlerinin BSTO değerlerinin değişimi paylaşılmıştır.

Şekil 6

Farklı Test Uzunluğu Koşullarına Göre Başarılı Sınıf Tahmini Oranları Grafiği



Şekil 6 incelendiğinde karar ağacına dayalı yöntemlerin tümünde test uzunlukları arttıkça BSTO değerlerinin arttığı, GDI, PWKL ve rastgele madde seçimi yöntemlerinde ise BSTO değerlerinin 12, 18 ve 24 maddelik test uzunluklarına göre anlamlı değişmediği gözlenmiştir. 12 maddelik test uzunluğuna sahip koşullarda, GDI'nin ardından en yüksek BSTO değerleri sırasıyla KA-2, KA-3, PWKL, KA-4 ve rastgele madde seçimi yöntemlerine aittir. 18 maddelik test uzunluğuna sahip koşullarda her ne kadar GDI madde seçimi yöntemi KA-3 ve KA-4 yöntemlerine yakın BSTO değerleri elde etse de en yüksek üç BSTO değerini karar ağacına dayalı yöntemler elde etmiştir. Karar ağacına dayalı farklı büyüklükte madde öbeklerinin kullanıldığı yöntemleri test uzunluğu açısından karşılaştırmak gerekirse, madde öbekleri büyüdükçe testteki madde sayısının azalmasından olumsuz anlamda daha fazla etkilendiği görülmektedir. 24 maddelik test uzunluğuna sahip koşullarda KA-2, KA-3

ve KA-4 yöntemlerinin oldukça yakın BSTO değerleri elde ettiği söylenebilir. Araştırmada kullanılan altı yöntemin tümünde 12, 18 ve 24 maddelik koşulların tamamında rastgele madde seçim yönteminin en düşük BSTO değerlerini elde ettiği saptanmıştır. Tablo 9'da ise 12, 18 ve 24 maddelik test uzunluğu koşullarında yürütülen analizlerden elde edilen ortalama başarılı beceri sınıflaması oranları paylaşılmıştır.

Tablo 9

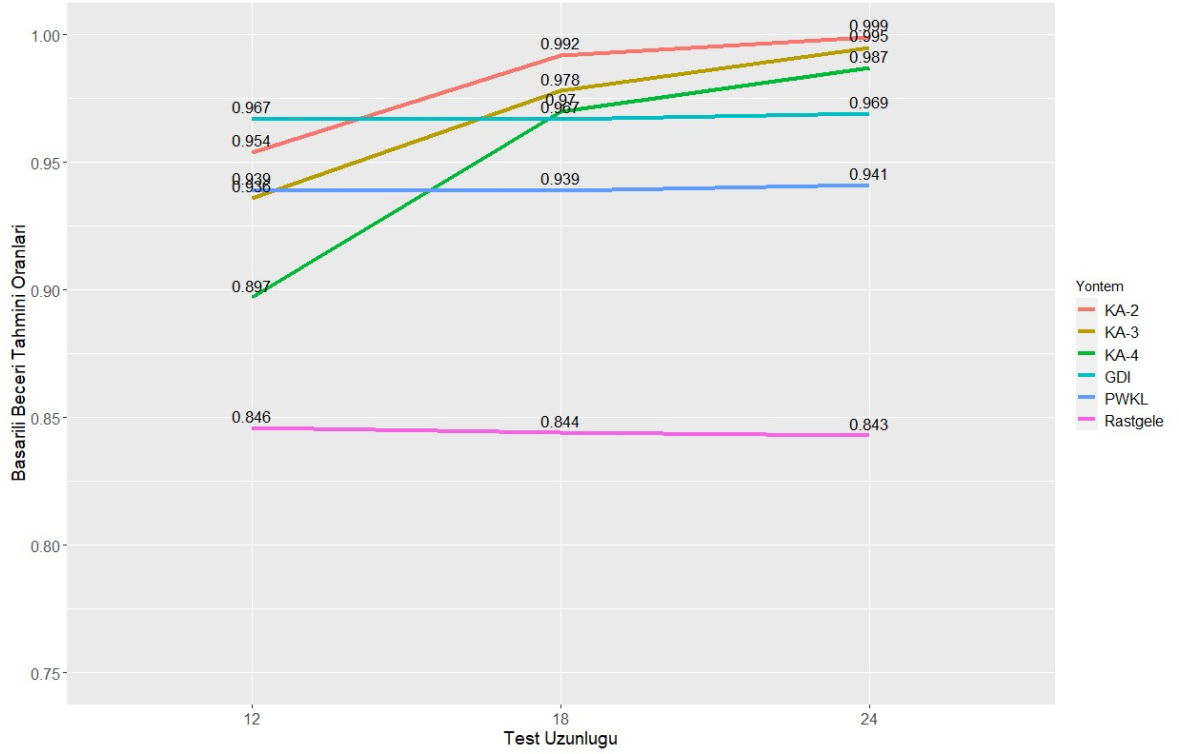
Farklı Test Uzunluklarına Göre Başarılı Beceri Tahmini Oranları

Test Uzunluğu	Karar Ağacı			GDI	PWKL	Rastgele
	2	3	4			
12	0,954	0,936	0,897	0,967	0,939	0,846
18	0,992	0,978	0,970	0,967	0,939	0,844
24	0,999	0,995	0,987	0,969	0,941	0,843

Tablo 9'a göre 12 maddelik test sonlandırma kuralının uygulandığı durumlarda BT-BBT'ye dayalı GDI ilerleme kuralının kullanıldığı analizlerin en yüksek başarılı beceri tahmini oranlarını elde ettiği tespit edilmiştir (0,957). Şekil 7'de farklı test uzunlukları koşulları altında madde seçimi yöntemlerine ait BBTO'ya etkisi gözlenebilir.

Şekil 7

Farklı Test Uzunluğu Koşullarına Göre Başarılı Beceri Tahmini Oranları Grafiği



Şekil 7'de yer alan farklı test uzunlukları koşullarına göre madde seçimi yöntemlerine ait BBTO değerleri Şekil 6'daki değerlerle beklenildiği gibi paralellik göstermektedir. BSTO değerlerine benzer şekilde 18 ve 24 maddelik koşullarda en yüksek üç BBTO değerini karar ağacı temelli yöntemler elde ettiği görülmüştür.

Karar ağacına dayalı farklı öbek büyüklüğünün kullanıldığı analizlerde ise 2 maddelik öbeklerin daha başarılı sonuçlar elde ettiği görülmektedir. Özellikle en kısa test uzunluğu olan 12 maddelik koşulda, en büyük öbeklerin bulunduğu 4 maddelik öbeklerin kullanıldığı uygulamaların göreceli kayda değer şekilde daha düşük başarılı beceri sınıflaması oranları elde ettiği saptanmıştır. Araştırmada kullanılan altı yöntemin tümünde 12, 18 ve 24 maddelik koşulların tamamında rastgele madde seçim yönteminin en düşük BBTO değerlerini elde ettiği saptanmıştır.

1.3. Alt Problemine İlişkin Bulgular

Bu alt problemde karar ağacı ve BT-BBT uygulamalarının başarılı sınıf tahmini oranlarının madde havuzu büyüklüklerine göre nasıl olduğu incelenmek istenmiştir. Bu amaç doğrultusunda iki farklı madde havuz büyüklüğüne göre (100, 200) elde edilen başarılı sınıf tahmini oranlarının ortalamaları Tablo 10'da sunulmuştur.

Tablo 10

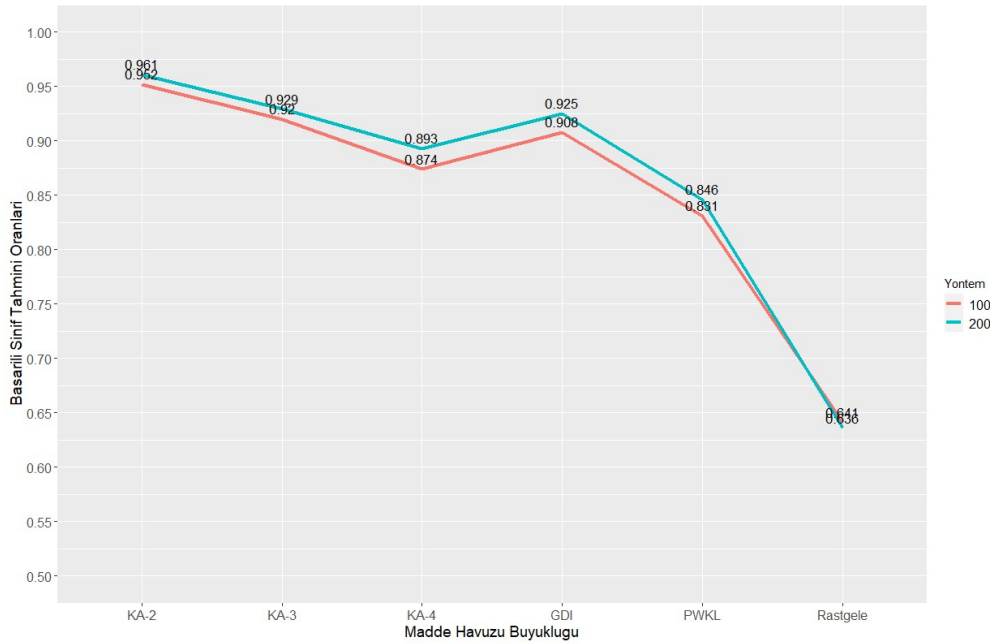
Farklı Madde Havuzu Büyüklüklerine Göre Başarılı Sınıf Tahmini Oranları

Havuz Büyüklüğü	Karar Ağacı (Öbek Büyüklüğü)			GDI	PWKL	Rastgele
	2	3	4			
100	0,952	0,920	0,874	0,908	0,831	0,641
200	0,961	0,929	0,893	0,925	0,846	0,636

Tablo 10 incelendiğinde rastgele madde seçim kuralına dayalı BT-BBT uygulaması dışında diğer 5 uygulama tipinde madde havuz büyüklüğünün 100'den 200'e çıkmasının başarılı sınıf tahmini oranlarını küçük oranda pozitif olarak etkilediği görünmüştür. Şekil 8 iki farklı madde havuzu büyüklüğü koşulu altında 6 madde seçimi yönteminin elde ettiği BSTO değerleri paylaşılmıştır.

Şekil 8

Madde Havuzu Büyüklüğüne Göre Başarılı Sınıf Tahmini Oranları



Şekil 8'e göre 100 ve 200 maddelik havuzların kullanıldığı simülasyon koşullarında elde edilen BSTO değerleri arasındaki farkların KA-2, KA-3, KA-4, GDI ve PWKL yöntemlerinde %1 ile %2 arasında değiştiği görülmüştür. Rastgele madde seçimi yönteminin ise madde havuzu büyüklüğünden anlamlı biçimde etkilenmediği gözlenmiştir. Ayrıca hem 100, hem de 200 maddelik havuzların kullanıldığı simülasyon koşullarının ortalamalarına göre en yüksek BSTO değerini KA-2 yöntemi elde etmiştir. 4 maddelik öbekler kullanılan karar ağacı uygulaması ve GDI madde seçimine dayanan BT-BBT uygulamasının ise diğer uygulamalara göre madde havuzu büyüklüğünden göreceli olarak daha fazla etkilendiği saptanmıştır. Tablo 11'de ise farklı havuz büyüklüğü koşullarına göre elde edilen ortalama başarılı beceri tahmini oranları sunulmuştur.

Tablo 11

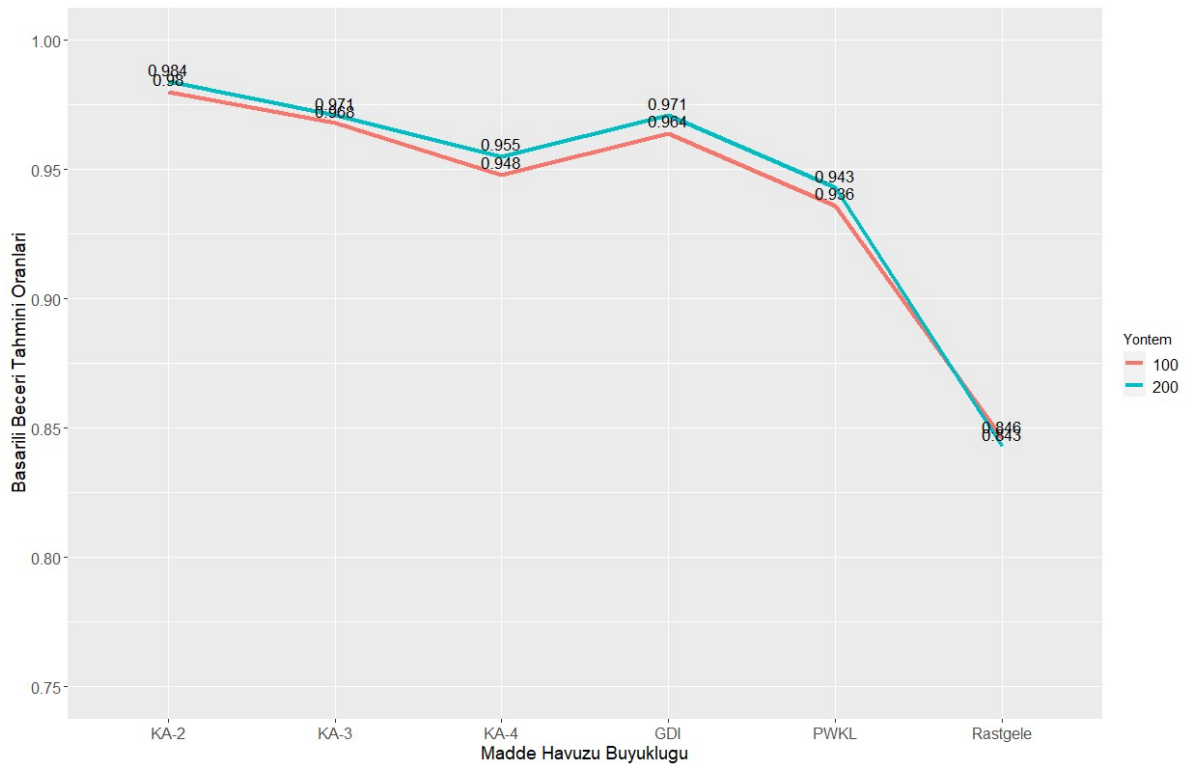
Farklı Madde Havuzu Büyüklüklerine Göre Başarılı Beceri Tahmini Oranları

Havuz Büyüklüğü	Karar Ağacı			GDI	PWKL	Rastgele
	2	3	4			
100	0,980	0,968	0,948	0,964	0,936	0,846
200	0,984	0,971	0,955	0,971	0,943	0,843

Tablo 11 incelendiğinde kullanılan 6 yöntemin 5'inde başarılı beceri tahmini oranlarının madde havuzunun büyümesinden az da olsa pozitif olarak etkilendiği görülmüştür. BT-BBT'ye dayalı rastgele madde seçimi yönteminde ise BBTO'nun madde havuzu büyüklüğünden anlamlı şekilde etkilenmediği gözlenmiştir.

Şekil 9

Madde Havuzu Büyüklüğüne Göre Başarılı Beceri Tahmini Oranları



Şekil 9'da Şekil 8'e paralel şekilde 100 ve 200 maddelik havuzların kullanıldığı simülasyon koşullarında elde edilen BBTO değerleri arasında farkların KA-2, KA-3, KA-4, GDI ve PWKL yöntemleri için %1'den daha küçük olduğu gözlenmiştir.

Kullanılan her iki havuz büyüklüğü koşulu için de en yüksek BBTO'yu KA-2 yönteminin elde ettiği görülmüştür. Bunun yanında KA-3 yöntemi de BBTO açısından her iki madde havuz büyüklüğü koşulu için de BT-BBT'ye dayalı tüm ilerleme kurallarından daha yüksek değerler elde etmiştir.

1.4 Alt Problemine İlişkin Bulgular

Bu alt problemde ise tüm madde ayırıcılık düzeyi, test uzunluğu ve madde havuzu büyüklüğü koşullarında karar ağacı temelli uygulamaların farklı madde öbeği büyüklüğü koşullarındaki başarılı sınıf tahmini oranlarını ortaya koymak amaçlanmıştır.

Bu kapsamda Tablo 4'ün en alt satırındaki 2, 3 ve 4 maddelik öbeklere ait tahmin oranları incelendiğinde 2 maddeden oluşan öbeklerin en yüksek başarılı sınıflama oranına sahipken 4 maddelik öbeklerin kullanıldığı uygulamalarda ise en düşük başarılı sınıflama oranlarının elde edildiği gözlenmiştir. BT-BBT temelli yaklaşımlar arasında ise GDI madde seçim yönteminin bireylerin örtük sınıflarını doğru tahmin etmede en başarılı yöntem olduğu söylenebilir. Bunu yanı sıra Tablo 5'in alt satırındaki ortalama değerler incelendiğinde en yüksek başarılı beceri tahmini oranının ortalamasının 2 maddelik öbekler kullanılan karar ağacına dayalı uygulamalardan elde edildiği söylenilebilir. BT-BBT temelli uygulamalar arasında ise GDI madde seçim yönteminin en başarılı sonuçları elde ettiği gözlenmiştir.

2. Alt Probleme İlişkin Bulgular

Bu çalışmanın ikinci alt probleminde gerçek veri ile yapılan uygulamada karar ağacı ve BT-BBT uygulamalarında başarılı beceri ve sınıf tahmini oranlarının karşılaştırılması amaçlanmıştır. Bu amaçla gerçek veri seti ve Q matrisi kullanılarak gerçekleştirilen uygulamalardan elde edilen başarılı sınıf ve yetenek tahmini oranları Tablo 12'de sunulmuştur.

Tablo 12*Gerçek Veri Kullanılarak Elde Edilen Başarılı Sınıf ve Beceri Tahmini Oranları*

Yöntem	Başarılı Sınıf Tahmini	Başarılı Yetenek Tahmini
KA-2	0,873	0,924
GDI	0,736	0,894
PWKL	0,691	0,881
Rastgele	0,615	0,847

Tablo 12 incelendiğinde Karar ağacına dayalı (2 maddelik öbek), GDI, PWKL ve Rastgele madde seçim yöntemleri kullanılarak aynı gerçek veri seti üzerinde yürütülen analizlere ait başarılı sınıf ve yetenek tahmini oranları görülmektedir. Buna göre en yüksek başarılı sınıf tahmini oranını karar ağacına dayalı uygulama 0,873 ile elde ederken, BT-BBT'ye dayalı yöntemler arasında GDI 0,736'lık oranla ikinci en yüksek başarı oranını elde etmiştir. En düşük başarılı sınıf tahmini oranını ise beklediği gibi rastgele madde seçimi yöntemi 0,615'lik oranla elde etmiştir. Başarılı yetenek tahmini oranlarında da yine karar ağacına dayalı uygulama 0,924'lük bir oran ile en başarılı uygulama olmuştur.

Gerçek veri ile yapılan uygulamada BT-BBT'ye dayalı yöntemler cdcR (Sorrel, Abad, 2021) paketi kullanılarak analiz edilmiş ve tüm uygulamalarda test uzunluğu 10 madde ile kısıtlanmıştır. Yani her bir birey tam olarak 10 maddelik testler almıştır. Buna karşın karar ağacı uygulamalarının gerçekleştirdiği "rpart" paketinde bireylerin karşılaşılabileceği madde sayılarını kesin olarak belirlemek teknik açıdan mümkün olmamıştır. Bu sebeple her bir birey için 10 maddelik testler kullanmak yerine toplam 3243 katılımcının uygulamaları maksimum 10-12 madde olacak şekilde kısıtlanmıştır. Bu kısıtlama oluşturulan ağaç tasarımındaki derinlik (depth) argümanını kullanarak gerçekleştirilmiştir. Gerçek uygulamada maddelerin bireylere 2'şer maddelik öbekler halinde yönlendirildiği düşünüldüğünde katılımcılar 4, 6, 8, 10 ve 12 maddelik testler ile karşılaşabilmektedir. Tablo

13'te 3243 katılımcıya ait farklı test uzunluklarını gösteren frekans ve yüzde çıktıları sunulmuştur.

Tablo 13

Karar Ağacı Uygulamasında Test Uzunluklarına Ait Frekans ve Yüzde Değerleri

Test Uzunluğu (Madde)	Frekans (Birey)	Yüzde
4	2238	69%
8	130	4%
10	713	22%
12	162	5%
Ortalama:5,88	3243	100%

Tablo 13 incelendiğinde katılımcıların %69'unun testlerinin 4 maddede (2 öbek) sonlandığı görülmektedir. Tüm katılımcılara ait test uzunluklarının ortalama değeri hesaplandığında ise gerçek veri ile gerçekleştirilen uygulamada karar ağacı uygulamalarının ortalama 5,88 maddede sonlandığı gözlenmiştir.

2.1. Alt Problemine İlişkin Bulgular

Araştırmanın bu alt probleminde karar ağacı ve BT-BBT temelli uygulamalarda maruz kalma oranlarının (item exposure rate) karşılaştırılması amaçlanmıştır. Bu kapsamda karar ağacı ve GDI madde seçim yöntemleri ile gerçek veri uygulamasından elde edilen maruz kalma oranları Tablo 14'te sunulmuştur.

Tablo 14*Karar Ağacı ve GDI Uygulamalarından Elde Edilen Maruz Kalma Oranları*

Madde No	Karar Ağacı	GDI
1	0,040	0,521
2	0,560	0,480
3	0,560	0,959
4	0,540	0,047
5	0,040	0,056
6	0,270	0,995
7	0,090	0,123
8	0,540	0,998
9	0,270	0,786
10	1,000	1,000
11	0,040	0,821
12	0,090	0,066
13	0,040	0,506
14	0,060	0,066
15	0,060	0,235
16	1,000	0,462
17	0,310	0,969
18	0,310	0,249
19	0	0,487
20	0	0,171
Ortalama	0,291	0,499

Tablo 14 incelendiğinde karar ağacı uygulamasında 2 maddenin maruz kalma oranının 1 olduğu yani her bir katılımcıya yönlendirildiği görülmüştür. 10 ve 16 numaralı maddeler kök nodunda yer aldığı için bu sonuç beklenmektedir. Ayrıca 19 ve 20 numaralı maddelerin hiçbir katılımcıya yönlendirilmediği, dolayısıyla maruz kalma oranlarının 0 olduğu tespit edilmiştir. BT-BBT'ye dayalı GDI madde seçim yöntemi ile gerçekleştirilen

uygulamada ise yüksek maruz kalma oranı olarak nitelendirilebilecek 7 madde olduğu görülmüştür (3, 6, 8, 9, 10, 11, 17). Bu 7 maddenin 5'i 0,95 üzerinde maruz kalma oranına sahiptir. Bunun yanı sıra 4 maddenin ise 0,10'un altında maruz kalma oranlarına sahip olduğu görülmektedir.

İki farklı uygulamaya ait maruz kalma oranlarının ortalamaları incelendiğinde ise karar ağacı yönteminin 0,291'lik, GDI uygulamasının ise 0,499'luk bir oran elde ettiği görülmektedir. Elde edilen bu oranlarda karar ağacı uygulamasında ortalama test uzunluğunun 5,88 madde iken BT-BBT uygulamalarında ise ortalama test uzunluğunun 10 madde olmasının etkisi olduğu söylenebilir.



Bölüm 5

Tartışma, Sonuç ve Öneriler

Bu bölümde araştırmanın bulgularından elde edilen sonuçlara ve önerilere yer verilmiştir.

Sonuçlar

Araştırmanın birinci alt probleminde karar ağacına dayalı BT-BBT uygulamalarından elde edilen Başarılı Sınıf Tahmini Oranlarının (BSTO) ve Başarılı Beceri Tahmini Oranlarının (BBTO) GDI, PWKL ve rastgele madde seçimi yöntemlerinin kullanıldığı geleneksel uygulamalar ile karşılaştırılarak performansların değerlendirilmesi amaçlanmıştır. Yöntemlerin elde ettiği tahmin başarıları oranları simülasyon koşullarına göre farklılık göstermektedir. Buna göre 2 maddelik öbeklerin kullanıldığı karar ağacı uygulamaları (KA-2) en yüksek BSTO ve BBTO değerlerini elde ettiği görülmüştür. KA-2'nin ardından sırasıyla KA-3, GDI, KA-4, PWKL ve Rastgele yöntemleri en yüksek BSTO ve BBTO değerlerini elde etmiştir.

Tüm yöntemlerin BSTO ve BBTO değerlerinin maddelerin ayırıcılık gücünün azalmasından beklenildiği gibi olumsuz etkilendiği gözlenmiştir. Düşük ve yüksek ayırıcılık gücü koşullarında elde edilen BSTO ve BBTO değerleri arasındaki farkın KA-2, KA-3 ve KA-4 diğer yöntemlere göre daha az olduğu tespit edilmiştir. Buna göre, karar ağacına dayalı yöntemlerin madde ayırıcılık gücündeki azalmaya diğer yöntemlere göre daha dayanıklı olduğu söylenebilir.

Karar ağacına dayalı yöntemlerin test uzunluğundan daha radikal bir biçimde etkilendiği gözlenmiştir. Bunun yanında karar ağacına dayalı 3 yöntemde kullanılan madde öbekleri küçüldükçe testin kısılmasına olan dayanıklılığın arttığı saptanmıştır. Buna göre KA-2'ye ait BSTO ve BBTO değerleri KA-3 ve KA-4'e göre olarak kısa test uzunluklarında daha yüksektir. Karar ağacı modelleri dallanma sayısı ve model derinliği arttıkça teorik yapıları gereği daha yüksek tahmin başarıları elde etmektedirler. Herhangi bir test uzunluğu

koşulunda madde öbeklerindeki madde sayısının artması karar ağacı modelindeki dallanma sayısını ve derinliğin azalması anlamına gelmektedir. Bu sebeple aynı test uzunluğu kullanıldığında küçük madde öbeklerinin kullanıldığı koşullar daha yüksek BSTO ve BBTO değerleri elde ettiği söylenebilir.

Karar ağacına dayalı yöntemler ve geleneksel madde seçimi yöntemlerinin madde havuzu büyüklüğüne dayanıklılık açısından 100 ve 200 maddelik havuz koşullarında önemli ölçüde farklılaşmadığı gözlenmiştir. Özellikle madde ayıricılığının yüksek olduğu durumlarda 100 ve 200 maddelik havuzların BSTO ve BBTO değerlerinin hemen hemen aynı olduğu dolayısıyla 100 maddelik havuzların bu koşullardaki uygulamalar için yeterli olduğu sonucuna ulaşılabilir.

Araştırmada kullanılan gerçek verilere göre en yüksek BSTO ve BBTO değerlerini KA-2'nin yakaladığı gözlenmiştir. Bu analizlerde kullanılan karar ağacı modelleri 2'şer maddelik öbekler ve 5 derinlik sınırı kullanılarak oluşturulmuştur. Bu durumda oluşturulan modelin maksimum 5 defa dallanmasına izin verilmektedir. Bu durum KA-2 uygulamasında bir öğrenciye maksimum 10 madde yönlendirilebileceği anlamına gelmektedir. Fakat, öğrencilerin birçoğu için testler 4, 6 ve 8 maddede (2,3 veya 4 öbek) sonlanmıştır. Bu frekanslar kullanılarak gerçek veri üzerinde yapılan KA-2 uygulamasında ortalama test uzunluğu 5,88 madde iken GDI, PWKL ve Rastgele yöntemleri söz konusu BSTO ve BBTO değerlerini sabit 10 maddelik testler ile elde etmiştir. Bu durumda test uzunluğu konusundaki dezavantajına rağmen KA-2'nin daha yüksek BSTO ve BBTO değerleri elde ettiği anlamına gelmektedir.

Gerçek veri kullanılarak yapılan KA-2 uygulamasında oluşturulan modelin kök nodunda bulunan iki madde karar ağacının yapısı gereği tüm öğrencilere uygulandığından maruz kalma oranları 1'dir. Yani bu iki madde tüm katılımcılara yönlendirilmiştir. Bu durumun test güvenliği açısından risk taşıdığı söylenilebilir. Fakat, kök nodundaki dallanmadan sonra havuzdaki diğer maddelere ait maruz kalma oranları hızla düşmektedir ve 0,60'ın üzerinde maruz kalma oranına sahip bir maddenin olmadığı gözlenmiştir. GDI

uygulamasında ise 1 oranına sahip bir madde bulunurken havuzda bulunan 20 maddenin 7 tanesinde bu oran 0,700'ün üzerinde olduğu görülmüştür. Tüm maddelerin maruz kalma oranlarının ortalamalarını karşılaştırdığımızda ise, KA-2 yöntemi için bu değer 0,291 iken GDI yönteminde 0,499 olarak hesaplanmıştır. Bu durumda KA-2 yönteminin maruz kalma oranları açısından, bir diğer deyişle madde havuzunun verimli kullanımı yönünden GDI yöntemine göre daha iyi sonuçlar elde ettiği söylenilebilir.

Araştırmacılara Öneriler

Araştırmada karar ağacına dayalı yöntemin farklı madde sayılarına sahip öbekler kullanılarak elde ettiği başarılı sınıflama oranlarını ortaya koymak için 2, 3 ve 4 maddelik öbekler ile simülasyon çalışmaları yürütülmüştür. Elde edilen sonuçlar öbeklerin küçülmesinin BSTO ve BBTO oranlarına pozitif yansıdığını göstermektedir. Bu durumda benzer bir çalışmanın her bir dallanmada madde öbekleri yerine tek maddenin kullanıldığı karar ağacı modelleri kullanılarak tekrarlanmasının anlamlı sonuçlar ortaya koyabileceği düşünülmektedir.

Araştırmada 100 ve 200 maddelik havuz büyüklükleri koşulları kullanılarak yöntemlerin havuz büyüklüğünden nasıl etkilendiği ve küçük madde havuzlarına karşı dayanıklılığı gözlenmeye çalışılmıştır. Elde edilen sonuçlar kullanılan madde türetme koşullarında hem 100 hem de 200 maddelik havuzların oldukça yeterli olduğunu ortaya koymuştur. Bu sebeple görece daha kısıtlı olan 100 maddelik koşulda dahi yüksek BSTO ve BBTO değerleri elde edildiği için farklı yöntemlerin küçük madde havuzlarına karşı dayanıklılıkları üzerine işlevsel çıktılar elde edilememiştir. Benzer bir çalışmanın daha küçük madde havuz koşulları ile tekrarlanmasının bu açıdan faydalı olacağı düşünülmektedir.

Uygulayıcılara Öneriler

Araştırmada elde edilen bulgulara göre karar ağacına dayalı yöntem ile elde edilen sonuçların madde sayısı, öbek büyüklüğü, madde ayıricılığı gibi koşullardan anlamlı

biçimde etkilendiği gözlenmiştir. Bu sebeple uygulayıcıların pilot uygulama sonrası elde edilecek parametreleri kullanarak çeşitli test uzunluğu ve öbek büyüklüğü koşulları ile simülasyonlar yürüterek kendi madde havuzları için ideal şartları araştırmasının sınıflama başarısını arttıracakı düşünülmektedir.

Özellikle görece kısa testlerin uygulanmasının hedeflendiği uygulamalarda büyük madde öbekleri oluşturmaktan kaçınılması tavsiye edilmektedir. Karar ağacı modellerinin en az 4 defa dallanmasına olanak verecek test uzunluğu-madde öbeği kombinasyonunun seçilmesinin sınıflama başarısını olumlu etkileyeceği söylenilebilir. Bu açıdan “hedeflenen test uzunluğu / madde öbeğinde madde sayısı” oranının 4 ya da daha büyük olması tavsiye edilebilir.

Kaynaklar

- Balamuta, J. J., Culpepper, S. A., & Hudson A. (2019) simcdm: Simulate Cognitive Diagnostic Model (CDM) Data. URL <https://cran.r-project.org/package=simcdm>.
- Başokçu, T. O. (2014). The cognitive diagnostic models for estimating students' ability and their applications. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 14(1), 1-32.
- Bartram, D., & Hambleton, R. K. (2006). Computer-based testing and the Internet. Issues and advantages. England: John Wiley and Sons.
- Chen, P., Xin, T., Wang, C., & Chang, H. H. (2012). Online calibration methods for the DINA Model with independent attributes in CD-CAT. *Psychometrika*, 201–222. <https://doi.org/10.1007/s11336-012-9255-7>
- Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika*, 74(4), 619-632. <https://doi.org/10.1007/s11336-009-9123-2>
- Cheng, Y. (2010). Improving cognitive diagnostic computerized adaptive testing by balancing attribute coverage: The modified maximum global discrimination index method. *Educational and Psychological Measurement*, 70(6), 902-913. <https://doi.org/10.1177/0013164410366693>
- Cheng, Y., & Chang, H.-H. (2009). The item selection method for cognitive diagnosis computerized adaptive testing. *Applied Psychological Measurement*, 33(8), 599–618. doi:10.1177/0146621608320526
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York, NY: Wiley.
- Davey, T. (2011). *A guide to computer adaptive testing systems*. Council of Chief State School Officers
- de Ayala, R.J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.

- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199.
- de la Torre, J., & Chiu, C. Y. (2010, April). General empirical method of Q-Matrix validation. Paper presented at the annual meeting of the National Council on Measurement in Education, Denver, CO.
- de la Torre, J., & Chiu, C. Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika*, 253-273. <https://doi.org/10.1007/s11336-015-9467-8>
- de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 333–353. <https://doi.org/10.1007/BF02295640>
- Delgado-Gomez, D., Laria, J. C., & Ruiz-Hernandez, D. (2019). Computerized adaptive test and decision trees: A unifying approach. *Expert Systems with Applications*, 117, 358-366.
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates.
- Guo, L., & Zheng, C. (2019). Termination rules for variable-length CD-CAT from the information theory perspective. *Frontiers in Psychology*, 10, 1122. <https://doi.org/10.3389/fpsyg.2019.01122>
- Hsu, C. L., Wang, W. C., & Chen, S. Y. (2013). Variable-length computerized adaptive testing based on cognitive diagnosis models. *Applied Psychological Measurement*, 37(7), 563-582. <https://doi.org/10.1177/0146621613488642>
- Hsu, C. L., & Wang, W.-C. (2015). Variable-length computerized adaptive testing using the higher order DINA model. *Journal of Educational Measurement*, 125-143. <https://doi.org/10.1111/jedm.12069>
- Huang, S. X. (1996, July). On content-balanced adaptive testing. In International Conference on Computer Aided Learning and Instruction in Science and Engineering (pp. 60-68). Berlin, Heidelberg: Springer Berlin Heidelberg.

- Kaplan, M., de la Torre, J., & Barrada, J. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied Psychological Measurement*, 167-188. <https://doi.org/10.1177/0146621614554650>
- Liu, H. Y., You, X. F., Wang, W. Y., Ding, S. L., & Chang, H. H. (2013). The development of computerized adaptive testing with cognitive diagnosis for an English achievement test in China. *Journal of Classification*, 30(2), 152-172. <https://doi.org/10.1007/s00357-013-9128-5>
- Liu, R., Huggins-Manley, A. C., & Bradshaw, L. (2016). The impact of Q-matrix designs on diagnostic classification accuracy in the presence of attribute hierarchies. *Educational and Psychological Measurement*, 77(2), 220-240. <https://doi.org/10.1177/0013164416645636>
- Ma, W., de la Torre, J. (2020). "GDINA: An R package for cognitive diagnosis modeling." *Journal of Statistical Software*, 93(14), 1–26. <https://doi.org/10.18637/jss.v093.i14>.
- Madison, M. J., & Bradshaw, L. P. (2015). The effects of Q-matrix design on classification accuracy in the log-linear cognitive diagnosis model. *Educational and Psychological Measurement*, 75(3), 491-511. <https://doi.org/10.1177/0013164414539162>
- McGlohen, M., & Chang, H. H. (2008). Combining computer adaptive testing technology with cognitively diagnostic assessment. *Behavior research methods*, 40, 808-821.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.
- Robitzsch, A., Kiefer, T., George A. C., & Uenlue, A. (2020). CDM: Cognitive Diagnosis Modeling. R package version 7.5-15, <https://CRAN.R-project.org/package=CDM>.
- Rodríguez-Cuadrado, J., Delgado-Gómez, D., Laria, J. C., & Rodríguez-Cuadrado, S. (2020). Merged Tree-CAT: A fast method for building precise computerized adaptive tests based on decision trees. *Expert Systems with Applications*, 143, 113066.
- Rubinstein, R. Y., & Kroese, D. P. (2016). *Simulation and the Monte Carlo method*. John Wiley & Sons.

- Shi, Q., Ma, W., Robitzsch, A., Sorrel, M. A., & Man, K. (2021). Cognitively diagnostic analysis using the G-DINA model in R. *Psych*, 3(4), 812-835.
- Silahtaroğlu, G. (2016). *Veri madenciliği* (3. Basım). Papatya Yayınları.
- Sorrel, M. A., Barrada, J., de la Torre, J., & Abad, F. (2020). Adapting cognitive diagnosis computerized adaptive testing item selection rules to traditional item response theory. *Plos One*, 15(1), 1-17. <https://doi.org/10.1371/journal.pone.0227196>
- Sorrel, M. A., Nájera, P., & Abad, F. J. (2021). cdcater: Cognitive Diagnostic Computerized Adaptive Testing. R package version 1.0.3. <https://CRAN.R-project.org/package=cdcatR>
- Stocking, M. L. (1994). Three practical issues for modern adaptive testing item pools 1. ETS Research Report Series, 1994(1), i-34.
- Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 51(3), 337-350. <https://doi.org/10.1111/1467-9876.00272>
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 345-354.
- Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In Paul D. Nichols, Susan F. Chipman, Robert L. Brennan (Ed.), *Cognitively Diagnostic Assessment* (pp. 327-359).
- Tatsuoka, K. K., & Tatsuoka, M. M. (1997). Computerized cognitive diagnostic adaptive testing: Effect on remedial instruction as empirical validation. *Journal of Educational Measurement*, 34(1), 3-20. <https://doi.org/10.1111/j.1745-3984.1997.tb00504.x>
- Therneau, T., & Atkinson, B. (2019). rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15. <https://CRAN.R-project.org/package=rpart>

- Wang, C. (2013). Mutual information item selection method in cognitive diagnostic computerized adaptive testing with short test length. *Educational and Psychological Measurement*, 73(6), 1017-1035. <https://doi.org/10.1177/0013164413498256>
- Wang, W., Song, L., Wang, T., Gao, P., & Xiong, J. (2020). A note on the relationship of the Shannon entropy procedure and the Jensen–Shannon divergence in cognitive diagnostic computerized adaptive testing. *SAGE Open*, 10(1), <https://doi.org/10.1177/2158244019899046>.
- Weiss, D. J., & Vale, C. D. (1987). Adaptive testing. *Applied Psychology*, 36(3-4), 249-262.
- Zheng, Y., Cheon, H., & Katz, C. M. (2020). Using machine learning methods to develop a short tree-based adaptive classification test: Case study with a high-dimensional item pool and imbalanced data. *Applied Psychological Measurement*, 44(7-8), 499-514.

EK-A: Araştırma Etik Komisyon İzin Muafiyeti Formu



Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Tez Çalışması/Araştırma Etik Komisyon İzin Muafiyeti Formu

F46

03/ 02/ 2023

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı Başkanlığına

Tez/Araştırma Başlığı

KARAR AĞACI ALGORİTMASININ BİLİŞSEL TANI MODELLERİNE DAYALI BİLGİSAYARDA
BİREYSELLEŞTİRİLMİŞ TESTLERDE KULLANIMI

Yukarıda başlığı/konusu verilen tez/araştırma çalışmam,

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır.
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir.
4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırmalar niteliğinde değildir.
5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir.

Çalışmada kullanacağım veriler:

() Kamusal erişime açık:

() Özel izin ve onaya tabi :

(x) Üretilmiş veri: R dili kullanılarak bilgisayar ortamında araştırmacı tarafından üretilmiş yapay veriler.

(x) Diğer: Daha önceden elde edilmiş, araştırmacı izni ile kullanılan, kişisel/kimlik verileri içermeyen çoktan seçmeli beşinci sınıf matematik sorularından elde edilmiş ikili veriler (1-0).

Yükseköğretim Kurumları Etik Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Hüseyin Yıldız

Araştırmacı Bilgileri

Adı Soyadı	Hüseyin Yıldız
Öğrenci ise No	N19141091
Ana Bilim Dalı	Eğitimde Ölçme ve Değerlendirme Anabilim Dalı
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	☐ Yüksek Lisans ☒ Doktora ☐ Bütünleşik Dr. ☐ Diğer

Danışman Görüşü ve Onayı*

*Tez ve tezden üretilen yayınlarda gerekli

(İmza)
(Danışmanın Ünvanı, Adı ve Soyadı)
Prof.Dr. Hülya Kelecioğlu

EK-B: Etik Beyanı

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- * tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- * görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- * başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- * atıfta bulunduğum eserlerin bütününe kaynak olarak gösterdiğimi,
- * kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- * bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

03/07/2024

Hüseyin YILDIZ

EK-C: Doktora Tez Çalışması Orijinallik Raporu

03/07/2024

HACETTEPE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü
Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı Başkanlığına,

Tez Başlığı: KARAR AĞACI ALGORİTMASININ BİLİŞSEL TANI MODELLERİNE DAYALI BİLGİSAYARDA BİREYSELLEŞTİRİLMİŞ TESTLERDE KULLANIMI

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
15/07 /2024	82	73877	11/07/2024	%3	2417244382

Uygulanan filtreler:

1. Kaynaklar hariç
2. Alıntılar dâhil
3. 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esaslarını inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Hüseyin YILDIZ

Öğrenci No.: N19141091

Ana Bilim Dalı: Eğitimde Ölçme ve Değerlendirme Anabilim Dalı

Programı: Eğitimde Ölçme Değerlendirme

Statüsü: ☐ Y.Lisans ☒ Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

Prof. Dr. Hülya KELECİOĞLU

EK-Ç: Dissertation Originality Report

03/07/2024

HACETTEPE UNIVERSITY
Graduate School of Educational Sciences
To The Department of Measurement and Evaluation in Education

Thesis Title: THE USE OF DECISION TREE ALGORITHM IN COMPUTERIZED ADAPTIVE TESTS BASED ON COGNITIVE DIAGNOSTIC MODELS

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
15/07 /2024	82	73877	11/07/2024	3%	2417244382

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Name Lastname: Hüseyin YILDIZ
Student No.: N19141091
Department: Eğitimde Ölçme ve Değerlendirme Anabilim Dalı
Program: Eğitimde Ölçme Değerlendirme
Status: ☐ Masters ☒ Ph.D. ☐ Integrated Ph.D.

ADVISOR APPROVAL

APPROVED
Prof. Dr. Hülya KELECİOĞLU

EK-D: Yayınlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- ☐ Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- ☐ Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren ... ay ertelenmiştir. ⁽²⁾
- ☐ Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

11 /03 /2024

Hüseyin YILDIZ

"*Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge*"

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tez erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6.2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internette paylaşılmaması durumunda 3. şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tez erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.

Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

*Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

