



2021 Yılı Ortaöğretim Kurumları Merkezi Yerleştirme Sınavında Öğrenci Başarılarının Veri Madenciliği Yöntemleriyle İncelenmesi

Fevzi KARACA^{1*} , Özlem ÇETİNKAYA BOZKURT² 

¹ Graduate Student, Burdur Mehmet Akif Ersoy University, Social Sciences Institute, Department of Management Information Systems, Burdur, Türkiye

² Prof. Dr., Burdur Mehmet Akif Ersoy University, Bucak Faculty of Business Administration, Department of Management and Organization, Burdur, Türkiye

Geliş Tarihi/Received: 04.04.2023
Kabul Tarihi/Accepted: 12.06.2023

Doi: 10.31200/makuubd.1277232
Araştırma Makalesi/Research Article

ÖZET

Yapılan bu çalışma ile ortaokul 8. sınıf öğrencilerinin Ortaöğretim Kurumları Merkezi Yerleştirme Sınavındaki (LGS) başarı durumları veri madenciliği yöntemleriyle incelenmiştir. Çalışmada kullanılan veriler Antalya ili Aksu ilçesinde öğrenim gören öğrencilere ait farklı özniteliklerden oluşan resmi verilerdir. Uygulamada veri madenciliğinin tanımlayıcı tekniklerinden birisi olan kümeleme (clustering) yönteminin farklı iki algoritması kullanılmıştır. Veri seti üzerinde İki adımlı kümeleme (two step clustering) ve K-ortalama (K-means) metotları kullanılmak suretiyle veri analizi yapılmıştır. Yapılan iki adımlı kümeleme işlemi sonucunda öğrencilerin %62'sinin 200-250 puan bandında yer aldığı tespit edilirken, k-ortalama ile yapılan kümeleme işleminde öğrenciler başarı durumlarına göre dört farklı kümeye ayrılmış ve her kümenin merkez puan değerleri ve tüm öğrencilerin bu merkezlere olan uzaklıklarına göre bir gruplandırma işlemi yapılmıştır. Yapılan analiz sonucunda ise öğrenci başarısında okul türü, cinsiyet ve okulun bulunduğu çevrenin başat faktörler olduğu görülmüştür.

Anahtar kelimeler: Eğitimde Veri Madenciliği, Öğrenci Başarı Analizi, İki Adımlı Kümeleme, K-Ortalama.

The Research of Student Achievements in Secondary Education Institutions Central Placement Exam in 2021 by using Data Mining Methods

ABSTRACT

The conducted study the success levels of the 8th-grade students in the Secondary Education Institutions Central Placement Exam (LGS) was examined by data mining methods. The data used in the study are official data consisting of different attributes belonging to the students studying in the Aksu district of Antalya province. In the study, two different clustering method algorithms, one of the descriptive techniques of data mining, were used. Data analysis was performed on the dataset by using two-step clustering and K-means methods. As a result of the two-step clustering process, while 62% of the students were found to be in the 200-250 score band, students were divided into four clusters according to their success levels in the clustering process with k-means and the center score values of each cluster and a grouping process according to all students' distance to these centers has been made. As a result of the analysis, it was seen that the school type, gender and the environment where the school is located were the dominant factors in student success.

Keywords: Data Mining in Education, Student Success Analysis, Two Step Clustering, K-Means.

1. GİRİŞ

Bilişim teknolojilerinin gelişimi ve kullanım alanlarının her geçen gün çoğalması büyük veri yığınlarını da beraberinde getirmektedir. Bu veri yığınları, işlenip ayrıştırılarak anlamlı ve işlevsel enformasyonlar haline gelebilmektedir. Veri madenciliği eldeki üstü kapalı, çok net olmayan verilerden, potansiyel olarak kullanışlı bilginin çıkartılmasıdır. Veriden enformasyona giden bu süreçte en çok kullanılan teknik veri madenciliğidir. Veri madenciliği uygulamaları bankacılık, pazarlama, e-ticaret, sigortacılık, tıp ve eğitim gibi birçok alan ve işkolunda kullanılmaktadır. Veri madenciliği teknikleri kullanılarak eğitim ve akademik başarı üzerine yapılan birçok çalışma bulunmaktadır.

Şen ve diğerleri (2012) çalışmalarında ortaöğretim yerleştirme sınavı sonuçlarını veri madenciliği yöntemleriyle tahminleme ve analiz etme yoluna gitmişlerdir. Sınav sonuçlarına ilişkin kapsamlı ve büyük bir veri seti kullanmışlardır. CRISP-DM metodolojisi ve k katmanlı çapraz doğrulama (k-fold cross validation) ile karar ağaçları (decision trees) teknikleri

kullanılarak başarı analizi yapılmış ve elde edilen sonuçların akademik başarının artırılması yönünde faydalı olacağı düşünülmüştür.

Bozkurt ve diğerleri (2015) Burdur Mehmet Akif Ersoy Üniversitesi Yönetim ve Bilişim Sistemleri Bölümü 3. ve 4. sınıf öğrencilerinin yönetim ve bilişim derslerindeki başarılarını karşılaştırmışlardır. 11 farklı öznelikten oluşan veri seti üzerinde Yapay Sinir Ağları (YSA) ve Uyarlamalı Sinir-Bulanık Sınıflayıcı (USBS) yöntemlerini kullanmışlardır.

Üçgün (2009) çalışmasında ortaöğretim kurumlarında okul yöneticileri tarafından kullanılan otomasyon yazılımları ve bu yazılımlara bağlı veritabanları üzerinde öğrenci başarısının analizi için veri madenciliği yöntemlerinden olan birliktelik kuramını kullanarak öğretim sürecinin düzenlenmesine yönelik sonuçları incelemiştir.

Shingari ve diğerleri (2017) tahmin edici veri madenciliği yöntemlerini kullanarak yükseköğrenim öğrencilerinin akademik performanslarını analiz etmişlerdir. Elde edilen bulgularla öğrencilerin akademik başarısının artacağını öngörmüşlerdir.

Kaur ve diğerleri (2015) her öğrencinin kişisel, psikolojik, sosyoekonomik ve diğer çevresel değişkenlerinin farklılık göstereceğini ve bu durumun öğrencilerin öğrenme hızlarını etkileyeceğini ifade etmişlerdir. Akranlarına göre daha yavaş öğrenen öğrencilerin başarı durumlarını incelemiştir. Çalışmalarında WEKA programında yapay sinir ağları, karar ağaçları ve K-en yakın komşu yöntemlerini kullanarak farklı sınıflandırma ve analizler yapmışlardır.

Abad ve López (2017) ilkokuldan seviyesinden lise seviyesine kadar farklı öğrencilerin matematik dersindeki akademik başarısını karar ağaçları ve yapay sinir ağları modellerini kullanarak araştırmışlardır.

Bu çalışmanın amacı 2021 yılında Ortaöğretim Kurumları Merkezi Yerleştirme Sınavına giren öğrencilerin okul türü, okul adı, okulun bulunduğu sosyokültürel çevre cinsiyet gibi değişkenler yönünden sınavda karşılaştıkları 6 farklı dersteki sorularda, toplam puanlarına ve yer aldıkları yüzdelerle ilişkilendirilerek başarı durumları incelenmektedir.

Çalışma sonucunda elde edilen bilgi ve bulgularla başarısızlık nedenlerinin belirlenmesi amaçlanmıştır. Böylelikle eğitim-öğretim faaliyetlerinde gerekli tedbirler alınarak ve doğru planlamalar yapılmak suretiyle eğitimcilere farklı modeller önerilecektir.

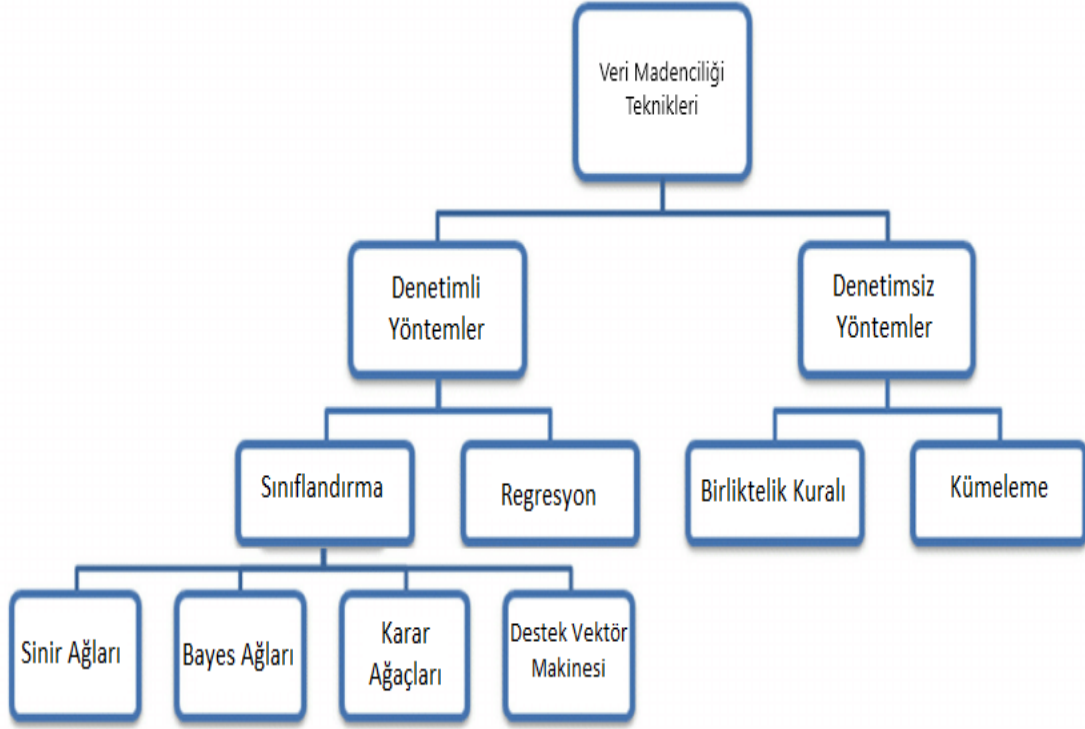
1.1. Veri Madenciliği

Veri madenciliği kavramı; ham halde bulunan veri yığınları arasından her türlü örgüt için değerli olan bilginin dinamik bir süreç ile tespiti işlemleridir. Elde edilen bilgiler ışığında karar verme süreçleri hızlandırılabilir, daha doğru kararlar verilebilir, seçme ya da eleme işlemleri yapılabilir ve gelecek için bir projeksiyon oluşturulabilir. Öte yandan veriler arasındaki benzerlikler, örüntüler ve bağıntılar tespit edilebilir. Böylece fayda sağlanacak alanın gelişimi hızlanır, üretim arttırılabilir, rekabet gücü artar ve örgüt kendini her zaman birkaç adım öne taşıyabilir (Çeşmeli vd., 2020, s.293).

Veriler arasındaki bilgi keşfi için istatistiksel yöntemlerin yetersiz kaldığı durumlarda veri madenciliği kullanılmaktadır. Çünkü veri sayısı ve öznitelik sayısı arttıkça verinin içinde saklı olan bilgiye ulaşmak için basit istatistiksel ve matematiksel yöntemler yetersiz kalmakta, veri madenciliği ön plana çıkmaktadır. Veri madenciliği sayesinde elde edilen bilgi ışığında karar vericiye yardım amaçlanmaktadır. Amaç asla karar vericinin yerine geçmek değil, aksine karar verme sürecini desteklemektir (Çeşmeli vd., 2020, s.294).

Bugün ticari işlemlerden, bilimsel verilere, uydu görüntülerine, eğitim alanına ve askeri istihbarata kadar elimizde tutabileceğimizden çok daha fazla bilgi vardır. Veri madenciliği, verileri farklı açılardan analiz etmek ve verilerden faydalı bilgiler elde etmek için etkili bir araç olmuştur. Artık birçok kuruluş, veri analizi için rekabetçi ortamlara başa çıkmak için veri madenciliğini bir araç olarak kullanmaya başlamıştır. Veri madenciliği, veri toplamak ve yönetmekten daha fazlasını içerir; aynı zamanda analiz ve tahmin içerir. Veri madenciliği hem özel sektörde hem de kamu sektöründe yaygınlaşmaktadır. Bankacılık, eğitim, sigortacılık, tıp ve perakendecilik gibi endüstriler, maliyeti azaltmak, araştırmayı geliştirmek, başarıyı ve satışları artırmak için yaygın olarak veri madenciliğini kullanırlar. Veri madenciliğinin temel amacı, minimum kullanıcı girdisi ve çabasıyla işlevsel veri modellerini otomatik olarak bulmaktır. Veri madenciliği, karar verme ve pazarın gelecekteki eğilimlerini tahmin etme becerisine sahip güçlü bir araçtır. Veri madenciliği araç ve teknikleri çeşitli alanlarda çeşitli şekillerde başarıyla uygulanabilmektedir (Dhote & Deshpande, 2016, s.211).

Veri madenciliği teknikleri, tahminleyici-denetimli yöntemler ve tanımlayıcı-denetimsiz yöntemler olmak üzere iki ana başlık altında açıklanabilir (Şekil 1).



Şekil 1. Veri madenciliği yöntemleri

Kaynak: Martín vd., 2014, s.612

Sonuçları önceden bilinen verilerden yola çıkılarak bir model geliştirilme aşaması ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi ve öngörülmesi amacı tahminleyici (öngörü) yöntemin tanımıdır diyebiliriz. Tanımlayıcı modellerde ise veri kümesindeki örüntülerin ve ilişkilerin bulunması amaçlanmaktadır.

1.2. Veri Seti ve Öznitelikler

Bu çalışmada Antalya Aksu İlçe Milli Eğitim Müdürlüğü tarafından ortaokullardan toplanan resmi veriler kullanılmıştır. Kullanılan verilerde 6698 sayılı Kişisel Verilerin Korunması Kanunu kapsamında hiçbir öğrencinin adı, soyadı veya Türkiye Cumhuriyeti vatandaşlık numarası gibi verileri kullanılmamıştır. Veriler 2021 yılı Haziran ayında yapılan Ortaöğretim Kurumları Merkezi Yerleştirme Sınavına giren 902 öğrencinin sınav sonuçlarından elde edilmiştir. Çalışmada kullanılan veriler öğrencilerin okul adı, okul türü, cinsiyet, dersler bazında doğru, yanlış ve boş bırakılan soru sayıları, yüzdelik dilim, toplam net, toplam doğru, toplam yanlış, toplam boş yanıt sayıları ve merkezi sistem sınav puanı olmak üzere toplam 27 adet öznitelikten oluşmaktadır. Bu öznitelikler Tablo 1’de yer almaktadır.

Tablo 1. Veri setinde yer alan öznitelikler ve anlamları

	Özelliği
Okul Türü	Okul türü (ortaokul veya imam hatip ortaokulu)
Cinsiyet	Kız veya Erkek
Türkçe Doğru	Türkçe dersinden doğru yanıtlanan soru sayısı
Türkçe Yanlış	Türkçe dersinden yanlış yanıtlanan soru sayısı
Türkçe Boş	Türkçe dersinden boş bırakılan soru sayısı
Matematik Doğru	Matematik dersinden doğru yanıtlanan soru sayısı
Matematik Yanlış	Matematik dersinden yanlış yanıtlanan soru sayısı
Matematik Boş	Matematik dersinden boş bırakılan soru sayısı
Fen Doğru	Fen Bilimleri dersinden doğru yanıtlanan soru sayısı
Fen Yanlış	Fen Bilimleri dersinden yanlış yanıtlanan soru sayısı
Fen Boş	Fen Bilimleri dersinden boş bırakılan soru sayısı
T.C. İnkılap Tarihi ve Atatürkçülük Doğru	T.C. İnkılap Tarihi ve Atatürkçülük dersinden doğru yanıtlanan soru sayısı
T.C. İnkılap Tarihi ve Atatürkçülük Yanlış	T.C. İnkılap Tarihi ve Atatürkçülük dersinden yanlış yanıtlanan soru sayısı
T.C. İnkılap Tarihi ve Atatürkçülük Boş	T.C. İnkılap Tarihi ve Atatürkçülük dersinden boş bırakılan soru sayısı
Din Doğru	Din Kültürü ve Ahlak Bilgisi dersinden doğru yanıtlanan soru sayısı
Din Yanlış	Din Kültürü ve Ahlak Bilgisi dersinden yanlış yanıtlanan soru sayısı
Din Boş	Din Kültürü ve Ahlak Bilgisi dersinden boş bırakılan soru sayısı
İngilizce Doğru	İngilizce dersinden doğru yanıtlanan soru sayısı
İngilizce Yanlış	İngilizce dersinden yanlış yanıtlanan soru sayısı
İngilizce Boş	İngilizce dersinden boş bırakılan soru sayısı
Sınav Puanı	Sınav Puanı
Yüzdelik Dilim	Türkiye Geneli Yüzdelik Dilimi
Toplam Doğru	Toplam Doğru Yanıtlanan Soru Sayısı
Toplam Yanlış	Toplam Yanlış Yanıtlanan Sayısı
Toplam Boş	Toplam Boş Bırakılan Soru Sayısı
Toplam Net	Toplam Net Sayısı

Araştırmada kullanılan veri setinde toplam 902 öğrenciye ait Tablo 1’de yer alan özniteliklerden oluşan veriler yer almaktadır. Öğrencilere ait en yüksek, en düşük ve ortalama değerler ile ilgili veriye ait standart sapma değerleri Tablo 2 de yer almaktadır.

Tablo 2. Veri setinde yer alan özniteliklere ilişkin istatistikî bilgiler

Öznitelik	En Küçük	En Büyük	Ortalama	Standart Sapma
TurkceDogru	0	20	7,59	3,974
TurkceYanlis	0	19	10,75	3,860
TurkceBos	0	13	1,66	2,231
MatematikDogru	0	19	3,49	2,477
MatematikYanlis	0	19	8,67	4,901
MatematikBos	0	20	7,84	6,104
FenDogru	0	19	6,56	3,790
FenYanlis	0	20	10,03	4,040
FenBos	0	20	3,40	3,741
InkDogru	0	10	4,36	2,513
InkYanlis	0	10	4,26	2,388
InkBos	0	10	1,39	1,929
DinDogru	0	10	5,45	2,526
DinYanlis	0	10	3,89	2,421
DinBos	0	9	0,66	1,239
IngDogru	0	10	3,86	3,053
IngYanlis	0	10	3,57	2,910
IngBos	0	10	2,55	3,351
SinavPuani	151,149	468,805	243,309	49,049434
YuzdelikDilim	0,42	99,95	60,5449	25,10178
ToplamDogru	6	84	31,31	13,716
ToplamYanlis	1	77	41,17	16,442
ToplamBos	30	95	47,52	13,932
ToplamNet	-12,67	83,67	17,5928	17,50097

2. YÖNTEM

Yapılan bu çalışmada tanımlayıcı veri madenciliği yöntemlerinden birisi olan “Kümeleme” metodunun iki farklı tekniği olan “Two-Step Cluster-İki adımlı Kümeleme” ve “K-Means Cluster- K-Ortalama Kümeleme” teknikleri kullanılmıştır. Verilerin analizinde SPSS programı kullanılmıştır. Veri seti üzerinde başarı analizi yapılırken öğrencilerin sınavda dersler bazında ve toplamda “net” olarak isimlendirilen ham başarı puanlarına bakılmaktadır. Bir derste “net” soru sayısı testte yer alan bir sorudaki seçenek sayısının bir eksiği kadar yanlış yanıtın, bir doğru yanıtı geçersiz hale getirmesi üzerinden hesaplanır. Ortaöğretim Kurumları Merkezi Yerleştirme Sınavında sorulan sorular 4 seçenekli olduğu için işaretlenen her 3 yanlış soru 1 doğru yanıtı etkisizleştirecektir.

2.1. İki Adımlı Kümeleme

İki Adımlı Küme Analizi tekniği, doğal yapıları ortaya çıkarmak için tasarlanmış bir keşif aracıdır. Bu yöntem ile çok büyük veri kümeleri işlenebilir. Bu teknik sayesinde hem kategorik hem de sürekli değişkenlere dayalı kümeler oluşturulabilir, küme sayısı otomatik olarak belirlenebilir ve büyük veri setleri ile çalışılabilir. İki aşamalı kümeleme analizinde ilk adım *ön kümeleme* ikinci adımda *kümeleme* olarak isimlendirilir. (Ceylan vd., 2017, s.478).

Ön kümeleme sürecinde veriler alt kümelere bölünerek ön kümeleme yapılır ve daha sonra bu alt kümeler ayrı veriler olarak değerlendirilir. Ele alınan verinin yeni bir alt kümemi oluşturacağı veya mevcut bir kümeye mi dâhil edileceği kararı uzaklık kriterlerine bağlı olarak hiyerarşik küme yöntemiyle yapılır (Rundle-Thiele vd., 2014, s.526).

Kümeleme sürecinde ise, ön kümeleme aşamasında oluşan alt kümeler, gerekli küme sayısına göre gruplara ayrılır. Daha sonra tüm kümeler karşılaştırılır ve aralarındaki mesafe en yakın olan küme çifti seçilir ve tek bir küme olarak birleştirilir. Birleşme sonrasında elde edilen yeni küme setleri karşılaştırılır ve en yakın küme çifti birleştirilir ve süreç tüm kümeler birleşene kadar yinelenir. İki aşamalı kümeleme tekniğinde aşağıdaki işlemler kullanılır (Tkaczynski, 2017).

Uzaklık Ölçüsü (Distance Measure)

Küme Sayısı (Number of Clusters)

Önem Ölçütü (Importance Measure)

2.1.1. Uzaklık ölçütü

İki aşamalı kümeleme analizi yapılırken her iki aşamada da uzaklık ölçüsüne gereksinim duyulur. Kullanılan veri setinde bir veya birden daha çok değişken kategorik olarak belirlenmişse, olasılığa bağlı uzaklığı temsil eden uzaklık ölçüsü kullanılır. Eğer tüm değişkenler sürekli değişken ise bu durumda Öklid uzaklığı kullanılır ve gözlemler küçük Öklid uzaklığına sahip kümede gruplanır (Bacher vd., 2004, s.4).

2.1.2. Küme sayısı

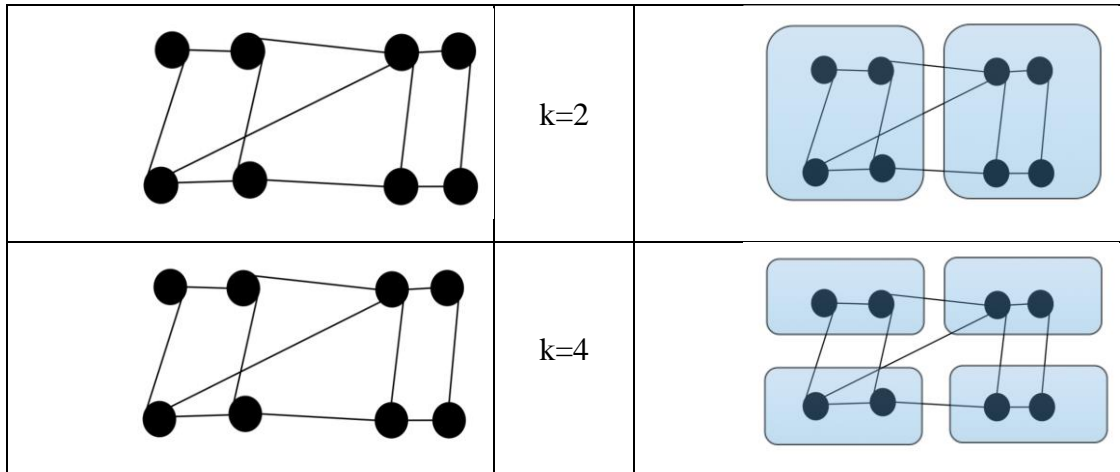
İki aşamalı kümeleme analizi, küme sayısının net olmadığı durumlarda kullanılır. Bu analiz yönteminde en uygun küme sayısı otomatik olarak belirlenir. Küme sayısı otomatik olarak belirlenirken Bayesian Information Criterion-BIC veya Akaike Information Criterion-AIC yöntemleri kullanılır (Ceylan vd., 2017, s.479).

2.1.3. Önem ölçütü

Kümeler oluşturulurken kullanılan değişkenlerin önemi 0-1 aralığında derecelendirilir. 0 değeri küme belirlemede en önemsiz değişkeni ve 1 değeri ise çok önemli değişkeni ifade eder (Ceylan vd., 2017, s.479).

2.2. K-Ortalama Kümeleme

K-Ortalama (K-Means) yöntemi kümeleme için kullanılan en eski ve popüler algoritmalarından bir tanesidir. Bu algoritma “k” tane kümenin ortalamasından oluşan bir değere göre kümelerin merkezlerini oluşturmaktır. K-Ortalama algoritmasının kullanılabilmesi için “k” sayısının önceden belirlenmiş olması gerekir. Aşağıdaki tabloda yer alan örnekte “k” sayısının 2 veya 4 olduğu durumlarda kümeleme işleminin nasıl ele alındığı gösterilmektedir (Fortunato, 2010, s.20).



Şekil 2. K-ortalama kümeleme

Kaynak: Şeker, 2015

K-Ortalama algoritması temelde 4 adımdan oluşur.

- Küme merkezlerinin belirlenmesi
- Merkez dışındaki değerlerin uzaklıklarına göre sınıflandırılması
- Yapılan sınıflandırmaya göre yeni merkezlerin belirlenmesi
- Kararlı bir yapı oluşturulana kadar 2. ve 3. adımların tekrarlanması

K-Ortalama algoritmasında amaç yapılan kümeleme işlemi sonunda ortaya çıkan kümelerin küme içi benzerliklerinin en fazla, kümeler arası benzerliklerinin ise en az seviyede olmasıdır (Şeker, 2015, s.32).

3. BULGULAR VE DEĞERLENDİRME

3.1. İki Adımlı Kümeleme Yöntemi Uygulaması Bulguları

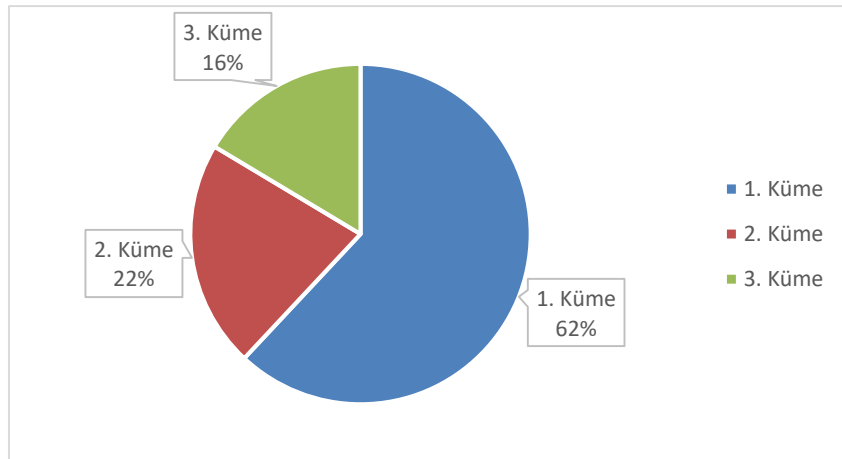
İki adımlı kümeleme uygulaması yapılırken nitel ve süreksiz değişkenler olarak “Okul Türü” ve “Cinsiyet” alanları nicel ve sürekli değişkenler olarak öğrencinin “Toplam Net Sayısı”, “Sınav Puanı” ve derslerdeki doğru sayılarını içeren alanlar referans alınmıştır.

“Okul Türü” ve “Sınav Puanı” arasındaki ilişkiyi betimlemek için yapılan iki adımlı kümeleme uygulamasına ilişkin sonuçlar Tablo 3’de görülmektedir. Modeli özetleyen şekilde algoritmanın iki adımlı kümeleme algoritması olduğu, iki değişken kullanıldığı, işlem sonucunda üç küme oluşturulduğu ve kümeleme işleminin üst düzey kalitede başarılı bir şekilde yapıldığı ifade edilmiştir.

Tablo 3. Okul türü ve sınav puanı bağlamında iki adımlı kümeleme özeti

Kullanılan Algoritma	İki Adımlı Kümeleme
Değişken Sayısı	2
Küme Sayısı	3
Uyum ve Ayrışma Bağlamında Küme Kalitesi (zayıf, yeterli, iyi)	İyi
Uyum ve Ayrışma Bağlamında Küme Kalitesi Katsayısı (1 üzerinden)	0,7

Şekil 3’de veri seti üzerinde yer alan 902 öğrencinin üç kümeye dağılımı gösterilmiştir. Birinci küme % 62, ikinci küme %21,6 ve üçüncü küme ise %16,4 oranlarında veri setini bölümlenmişlerdir. En büyük kümenin en küçük kümeye oranı 3,78 olarak elde edilmiştir. Kümeleme yapılırken kullanılan değişkenlerin önem düzeyleri 0 ile 1 arasında değer almaktadır. Buradaki değer 1 değerine yaklaştığı ölçüde tahminleyici önemi (predictor importance) artacaktır. Kullanılan iki değişkenden “Okul Türü” değişkeninin önem düzeyi 1, Sınav Puanı değişkeninin önem düzeyi ise 0,91 olarak belirlenmiştir.

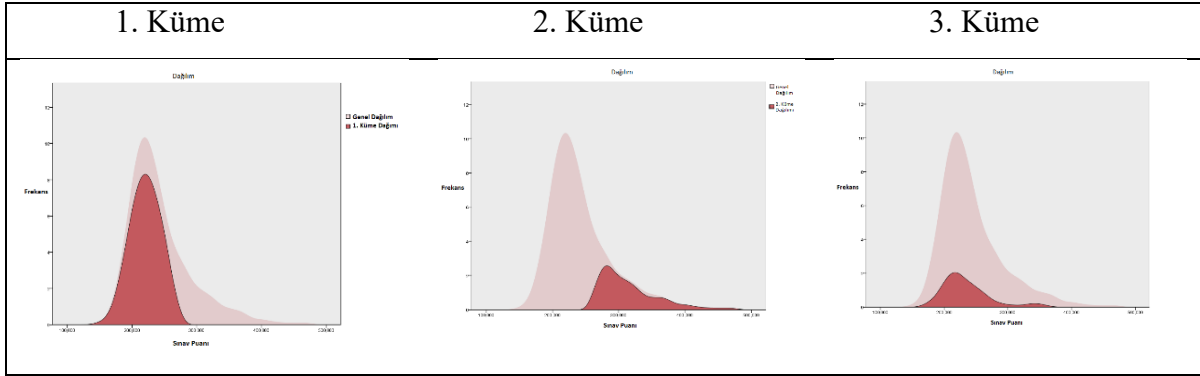


Şekil 3. Verilerin kümelere dağılımı

Her üç kümeye ait sınav puanı dağılım grafikleri ve sınav puanlarına ait frekanslar Tablo 4’de gösterilmektedir. 559 öğrencinin birinci kümede yer aldığı, puan ortalamalarının 220,740, okul türlerinin “ortaokul”, birinci çeyreğin 204,980, ikinci çeyreğin (medyan) 220,500 ve üçüncü çeyreğin 237,530 sınav puanı değerlerinden oluştuğu görülmüştür. En başarılı öğrencilerin ikinci kümede toplandığı ve bu kümede 195 öğrencinin yer aldığı görülmektedir. Bu öğrencilerin puan ortalamasının 315,070, okul türlerinin “ortaokul”, birinci çeyreğin 280,890, ikinci çeyreğin (medyan) 303,620 ve üçüncü çeyreğin 334,000 sınav puan değerinden oluştuğu gözlemlenmiştir. Üçüncü ve son kümede yer alan öğrenci sayısının 148, puan ortalamalarının 234,030, okul türlerinin “imam hatip ortaokulu”, birinci çeyreğin 209,850, ikinci çeyreğin (medyan) 225,900 ve üçüncü çeyreğin ise 250,220 puan olduğu belirlenmiştir.

Öğrencilere ait puan dağılımları 300,000 puana kadar normal dağılım gösterirken, 300 ile 500 puan arasında sağa çarpık bir dağılım söz konusudur. En çok öğrencinin ise 200 ile 250 puan aralığında yer aldığı gözlemlenmiş ve sınava giren tüm öğrencilerin sınav puanı ortalaması 243,309 olarak hesaplanmıştır. Üç küme şeklinde gruplandırılan öğrencilerin puan dağılım grafikleri Tablo 4’de gösterilmiştir.

Tablo 4. Kümelerin puan dağılımları

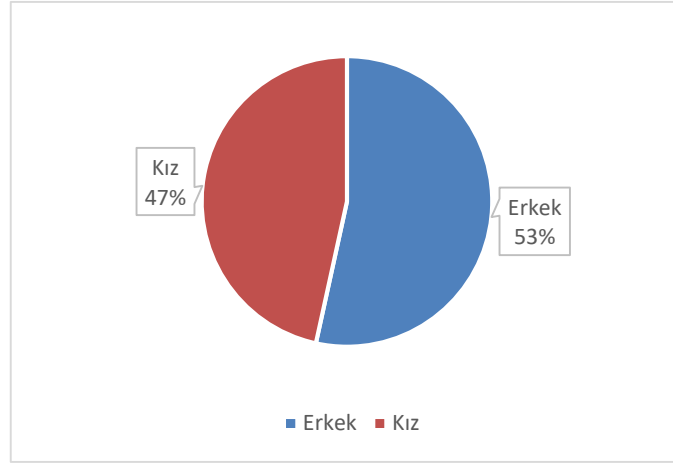


“Cinsiyet” ve “Toplam Net Soru Sayısı” arasındaki bağıntıyı incelemek için yapılan ikinci iki adımlı kümeleme uygulamasında elde edilen sonuçlar Tablo 5’de yer alan model özetinde görülmektedir. Algoritmanın iki adımlı kümeleme algoritması olduğu, iki değişken kullanıldığı, işlem sonucunda iki küme oluşturulduğu ve kümeleme işleminin üst düzeyde başarılı bir şekilde sonuçlandığı gösterilmiştir.

Tablo 5. Cinsiyet ve toplam net sayısı bağlamında iki adımlı kümeleme özeti

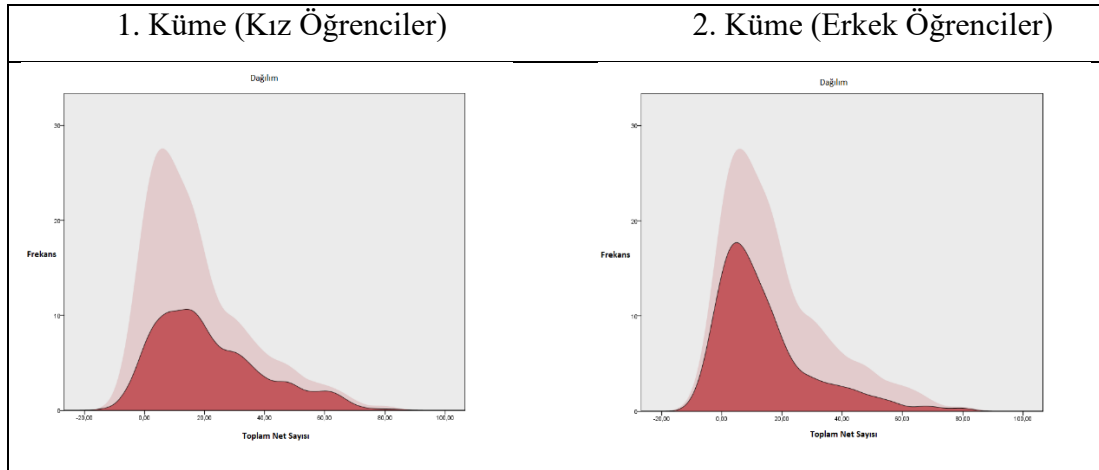
Kullanılan Algoritma	İki Adımlı Kümeleme
Değişken Sayısı	2
Küme Sayısı	2
Uyum ve Ayrışma Bağlamında Küme Kalitesi (zayıf, yeterli, iyi)	İyi
Uyum ve Ayrışma Bağlamında Küme Kalitesi Katsayısı (1 üzerinden)	0,8

Şekil 4’de veri setinde yer alan öğrencilerin oluşturulan iki kümeye dağılımları ve oransal durumları gösterilmektedir. Buna göre sınava giren erkek öğrenci sayısı 482, kız öğrenci sayısı da 420’dir. Seçilen örneklemin %53,4’ü erkek, %46,6’sı ise kız öğrencilerden oluşmaktadır. Erkek öğrenci sayısının kız öğrenci sayısına oranı 1,15 olarak belirlenmiştir. Kümeleme işleminde kullanılan başat değişken “cinsiyet” alanı olup, önem düzeyi 1’dir. Öğrencilerin cinsiyete göre oransal dağılımları Şekil 5’de yer alan grafikte gösterilmiştir.



Şekil 4. Verilerin cinsiyete bağlı kümelere dağılımı

Birinci küme 420 öğrenciden oluşmakta ve bu öğrencilerin tamamı kız öğrencilerdir. Kız öğrencilerin toplam net ortalaması 21,63’tür. Bu öğrenci grubuna ait birinci çeyrek 7,78, ikinci çeyrek (medyan) 17,83 ve üçüncü çeyrek ise 31,95 değerinden oluşmaktadır. İkinci kümede yer alan 482 öğrencinin tamamı erkektir. Bu kümedeki öğrencilerin net ortalaması 14,07’dir. Erkek öğrencilerin birinci çeyreği 3,06, ikinci çeyreği (medyan) 10,02 ve üçüncü çeyreği ise 19,22 olarak gözlemlenmiştir. Öğrencilerin cinsiyete bağlı toplam net sayıları Tablo 6’da yer alan grafiklerde gösterilmiştir.

Tablo 6. Öğrencilerin cinsiyet ve toplam net sayıları ilişki grafiği

Toplam net sayılarına ilişkin ortalamalar göz önüne alındığında kız öğrencilerin erkek öğrencilerden daha başarılı olduğu belirlenmiştir. Bunun yanında toplamda 75 net sayısının üzerine çıkan öğrencilerin de erkek öğrenciler olduğu belirlenmiştir. Erkek öğrencilerinden 0 net sayısının altında yer alan öğrenci sayısının kız öğrencilere göre daha fazla olduğu, kız öğrencilerin 15-20 net sayısı aralığında, erkek öğrencilerin ise 10-15 net aralığında yoğunlaştığı gözlemlenmiştir. Kız öğrencilerin net sayısı kararlı ve dengeli bir dağılım göstermemektedir. Erkek öğrenciler de durum biraz daha farklı olup, 30 net sayısı bandına kadar dengeli bir dağılım oluşmaktayken, 30 net üzeri dağılımda grafik sağa çarpık bir eğilim göstermektedir.

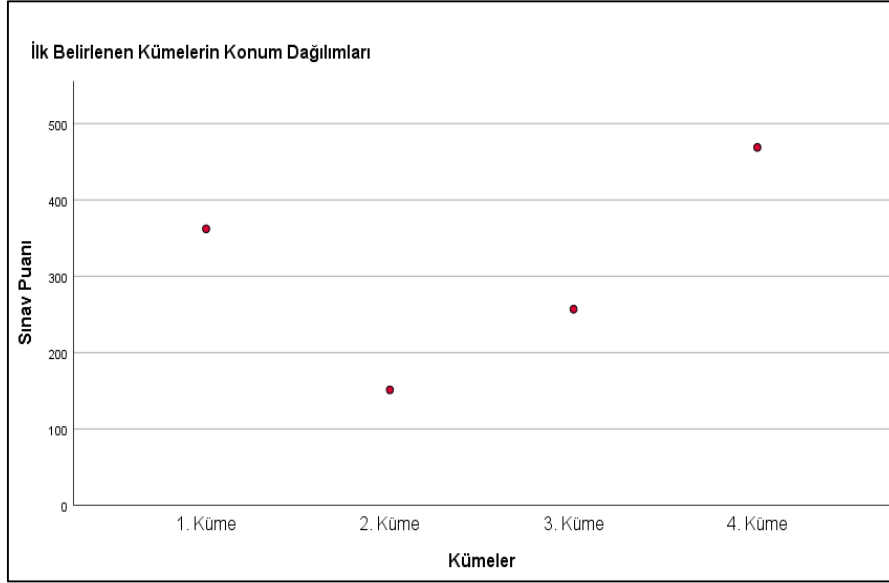
3.2. K-Ortalama Yöntemi Uygulaması Bulguları

K-Ortalama yöntemiyle yapılan kümeleme birinci kümeleme işleminde okul türü ve sınav puanı esas alınmıştır. Burada okul türü bağımlı, sınav puanı ise bağımsız değişkendir. Bir başka ifadeyle okul türü ve sınav puanı arasındaki ilişki incelenmiştir. Küme sayısı, iki adımlı kümeleme işleminde otomatik olarak belirlenirken burada K-Ortalama algoritmasında küme sayısı yani “k” değeri 4 olarak belirlenmiştir. Alanda yapılan çalışmalar incelendiğinde “k” merkez sayısı değerinin hangi doğrultuda veya hangi veriler ışığında belirlendiğine dair çok fazla açıklama veya dayanak bulunmamaktadır (Pham vd., 2005, s.103). “k” değerinin belirlenmesi genellikle objektif norm ve kriterler dışında araştırmacıların deneme yanılgıları neticesinde öznel bir seçim olmaktadır (İlkin vd., 2020, s.186). Bu bağlamda sınava giren öğrencilerin 0 ile 500 puan aralığında bir skor elde edeceği göz önüne alındığında, bölütlemenin 4 farklı merkez ekseninde yapılmasının yeterli olacağı düşünülmüştür. Yapılan kümeleme işleminde ilk adımda 4 tane küme merkezi belirlenmiştir. Veri seti üzerinde yer alan değerler oluşturulan bu 4 küme merkezi etrafında gruplanmıştır. Tablo 7’de ilk yapılan işlem sonrasında

oluşan 4 farklı küme merkez değerleri gösterilmiştir. Şekil 5’de ise kümelerin sınav puanı değişkenine bağlı olarak oluşan ilk konumları gösterilmiştir.

Tablo 7. İlk kümeleme işlemi sonrasında küme merkezi değerleri

	Küme			
	1	2	3	4
Sınav Puanı	362,137	151,149	256,869	468,805

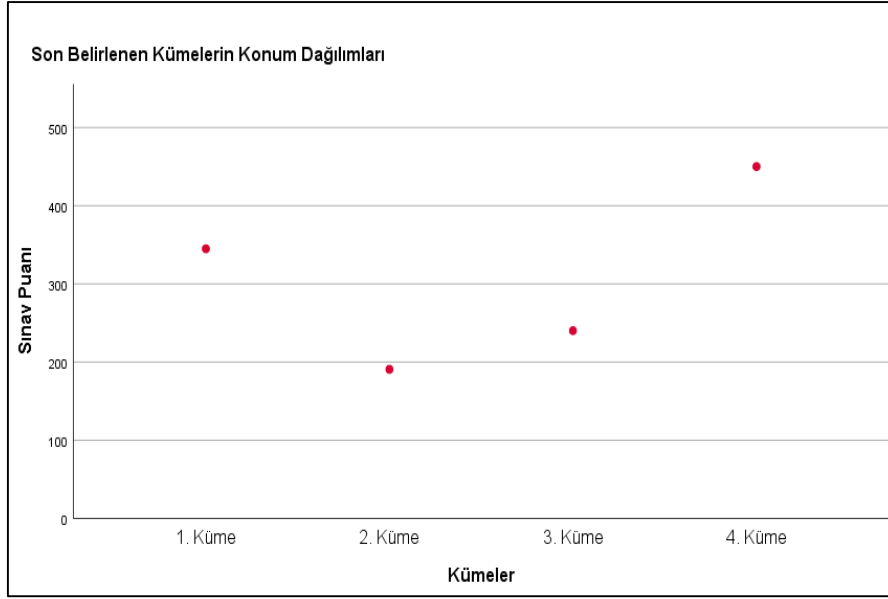


Şekil 5. Belirlenen kümelerin ilk konum dağılımları

Merkez dışında kalan değerler 4 farklı merkez noktaya olan uzaklıklarına göre sınıflandırılmıştır. Bu işlem eldeki yapının kararlı hale gelmesi için 10 defa tekrarlanmıştır. Yapılan 10 tekrarlama işlemi (iterasyon) sonrasında oluşan yeni küme merkezleri ve bu merkezlere ait değerler Tablo 8’de gösterilmiştir. Kararlı hale gelen küme merkezlerinin sınav puanına bağlı olarak oluşan son konumları da Şekil 6’da yer alan grafikte gösterilmiştir.

Tablo 8. Son kümeleme işlemi sonrasında küme merkezi değerleri

	Küme			
	1	2	3	4
Sınav Puanı	344,864	190,741	240,184	450,014



Şekil 6. Belirlenen kümelerin son konum dağılımları

Veri seti üzerindeki sınav puanı değeri Tablo 8’de yer alan küme merkezlerine olan uzaklıklarına göre gruplara ayrılmıştır. Herhangi bir değerin küme merkezine olan uzaklığı hesaplanırken en yaygın kullanılan doğrusal uzaklık hesaplama yöntemlerinden biri olan “Öklid Uzaklığı” kullanılmıştır. K-Ortalama yöntemiyle yapılan kümeleme analizinde kararlı bir yapı oluşturulduktan sonra öğrencilerin başarı durumlarına göre sırasıyla 4. kümede, 1. kümede, 3. kümede ve 2. kümede yer aldıkları görülmüştür.

En başarılı olan öğrencileri içeren 4. küme aynı zamanda elaman sayısı en az olan kümedir. Veri seti üzerinde yer alan 902 öğrenciden sadece 6 tanesi 4. kümede yer alabilmektedir. Bu 6 öğrencinin tamamının okul türünün “ortaokul” ve okullarının ilçe merkezinde yer alan okullar olduğu gözlemlenmiştir. Öğrencilerin sınav puanları ve küme merkezlerine olan uzaklıkları Tablo 9’da gösterilmiştir.

Tablo 9. 4. Kümede yer alan öğrenci bilgileri

Sıra Numarası	Okul Türü	Ait Olduğu Küme	Küme Merkezine Olan Uzaklık	Sınav Puanı
40	Ortaokul	4	0	468,805
41	Ortaokul	4	1,0102	467,795
42	Ortaokul	4	37,8558	430,949
286	Ortaokul	4	30,2907	438,514
327	Ortaokul	4	10,6835	458,122
328	Ortaokul	4	32,9054	435,900

Diğer kümeler incelendiğinde 1. Kümede 90 öğrenci, 3. Kümede 647 öğrenci ve son küme olan 2. Kümede 159 öğrencinin yer aldığı görülmüştür.

4. TARTIŞMA VE ÖNERİLER

Günümüzde veri madenciliği yöntemleri sadece iş ve finans dünyasında kullanılmamaktadır. İlerleyen dönemde eğitim alanında da yoğun olarak kullanılacaktır. Öğrencilerin akademik ve sosyal başarılarının artırılması 21. yüzyıl eğitimcilerinin en büyük hedeflerinden bir tanesi olup, üzerinde yoğun olarak çalışılan bir konudur. Veri madenciliği araçlarıyla yapılan analiz ve incelemeler neticesinde elde edilen veriler ile günümüz eğitim uygulamalarının akademik, bilişsel, duyuşsal, sosyolojik ve yönetsel boyutları farklı biçimlerde değerlendirilebilecektir. Bu bağlamda eğitsel içerik geliştirme süreçlerinde ve pedagojik kararlar almada etkili olacaktır.

Bu çalışmanın amacı ortaokul 8. sınıfta öğrenim gören öğrencilerin yılsonunda girdikleri ortaöğretim kurumları merkezi yerleştirme sınavındaki başarılarını incelemektir. Çalışmada geleneksel istatistikî yöntemler yerine “iki adımlı kümeleme” ve “k-ortalama” teknikleri kullanılmıştır. Bu kümeleme teknikleri veri madenciliği ve veri analizinin merkezinde yer alarak yoğunlukla kullanılmaktadır. Kümeleme analizi ile karmaşık veri toplulukları belli değişkenler referans alınarak daha anlamlı hale getirilmesidir. Elde edilen sonuçlar sayesinde karar verme süreçlerine destek sağlanır. Kümeleme sayesinde büyük veri yığınları içerisinde gizli kalmış ama kullanışlı olabilecek bilgiler ortaya çıkarılabilir. Kümeleme algoritmaları tarafından etkin şekilde ele alınan veriler daha anlaşılır ve işlevsel hale getirilebilir.

Yapılan çalışmada öğrencilere ait cinsiyet, okul türü ve okulun bulunduğu bölge gibi değişkenler göz önüne alınarak başarı durumları incelenmiştir. Yapılan çalışma neticesinde farklı başarı durumlarına göre ayrıştırılan öğrenci sayıları tespit edilmiştir. Sınavla öğrenci alan üst düzey nitelikli okullara kayıt yaptırmaya hak kazanan öğrenci sayısının, genel öğrenci sayısı dikkate alındığında yeterli düzeyde olmadığı görülmüştür. Öğrencilerin okul türleri incelendiğinde ortaokullarda öğrenim gören öğrencilerin imam hatip ortaokullarında öğrenim gören öğrencilere göre daha başarılı olduğu belirlenmiştir. İlçe merkezinde yer alan okullarda öğrenim gören öğrencilerin kırsal bölgelerde yer alan okullarda öğrenim gören öğrencilere oranla daha iyi sınav sonuçları aldığı gözlemlenmiştir. Genel öğrenci başarısında cinsiyet faktörü ve puan ortalamaları göz önüne alındığında kız öğrencilerin erkek öğrencilerden daha

başarılı olduğu görülmesine karşın genel başarı sıralamasında üst sıralarda yer alan öğrenciler içerisinde erkek öğrencilerin de çoğunlukta olduğu bir tablo ortaya çıkmıştır.

Sonuç olarak; Ortaöğretim Kurumları Merkezi Yerleştirme Sınavı her öğrencinin hayatı boyunca sadece bir defa girebileceği bir sınav olup nitelikli bir ortaöğretim kurumuna yerleşebilmesi açısından son derece önemlidir. Bu nedenle istenilen başarı seviyesine ulaşamayan öğrencilerin derslerdeki kazanım eksiklerinin yıl içerisinde yapılacak çalışmalarla tespit edilmesi ve okulları tarafından gerekli önlemlerin alınması son derece önemlidir. Bu çalışmada genel bir başarı analizi yapılmıştır. İlerleyen süreçte Ortaöğretim Kurumları Merkezi Yerleştirme Sınavında soru sorulan alanlara kaynaklık eden altı farklı derse ait daha detaylı ve kapsamlı incelemeler yapılabilir. Yapılan bu çalışmalarla başarısızlık nedenleri ortaya çıkarılabilir, doğru yorumlar yapılarak başarıyı artırma yönünde çözümler önerilebilir.

ARAŞTIRMACILARIN KATKI ORANI BEYANI

Araştırmacıların çalışmaya katkı oranları eşittir.

ÇATIŞMA BEYANI

Çalışma kapsamında herhangi bir kurum veya kişi ile çıkar çatışması bulunmamaktadır.

REFERENCES / KAYNAKLAR

- Abad, F., & López, A. (2017). Data-mining techniques in detecting factors linked to academic achievement. *School Effectiveness and School Improvement*, 28(1), 39-55. doi: <https://doi.org/10.1080/09243453.2016.1235591>
- Bacher, J., Wenzig, K., & Vogler, M. (2004). SPSS twostep cluster-a first evaluation. *Lehrstuhl für Soziologie - Arbeits- und Diskussionspapiere*.
- Bozkurt, Ö., Kalkan, A., Pençe, İ., & Çeşmeli, M. (2015). Yönetim bilişim sistemleri öğrencilerinin yönetim ve bilişim derslerindeki başarılarının veri madenciliği yöntemleri ile incelenmesi. *Yönetim Bilişim Sistemleri Dergisi*, 2, 36-47.
- Ceylan, Z., Gürsev, S., & Bulkan, S. (2017). İki aşamalı kümeleme analizi ile bireysel emeklilik sektöründe müşteri profilinin değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10(4), 475-485.
- Çeşmeli M., Bozkurt Ö., & Pençe İ. (2020). *Yönetim bilişim sistemleri*. Ankara: Nobel.
- Dhote, R. A., & Deshpande, S. P. (2016). Data mining with cloud computing: An overview. *International Journal of Advanced Research in Computer Engineering & Technology*, 5(1).
- Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3-5), 75-174.

İlkin, S., Aytar, O., Gençtürk, T. H., & Şahin, S. (2020). Dermskopik görüntülerde lezyon bölütleme işlemlerinde k-ortalama kümeleme algoritmasının kullanımı. *Gazi University Journal of Science Part C: Design and Technology*, 8(1), 182-191.

Kaura, P., Singhb, M., & Josanc, G. S. (2015). Classification and prediction based data mining algorithms to predict slow learners in education sector. *Science Direct*, 57, 500-508.

Martín, L., Baena, L., Garach, L., López, G., & De Oña, J. (2014). Using data mining techniques to road safety improvement in Spanish roads. *Procedia-social and Behavioral Sciences*, 160, 607-614.

Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1), 103-119.

Rundle-Thiele, S., Kubacki, K., Tkaczynski, A., & Parkinson, J. (2015). Using two-step cluster analysis to identify homogeneous physical activity groups. *Marketing Intelligence & Planning*, 33(4), 522-537.

Shingari, I., Kumar, D., & Khetan, M. (2017). A review of applications of data mining techniquesfor prediction of students' performance in higher education. *Journal of Statistics and Management Systems*, 20, 713-722. doi:<https://doi.org/10.1080/09720510.2017.1395191>

Şeker, S. E. (2015). Sosyal ağlarda veri madenciliği (data mining on social networks). *YBS Ansiklopedi*, 2(2), 30-39.

Şen, B., Ucar, E., & Delen, D., (2012). Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*, 39(10), 9468-9476.

Tkaczynski, A. (2017). Segmentation using two-step cluster analysis. *Segmentation in Social Marketing: Process, Methods and Application*, 109-125.

Üçgün, K. (2009). *Ortaöğretim okulları için öğrenci otomasyonu tasarımı ve öğrenci verileri üzerine veri madenciliği uygulamaları* (Basılmamış yüksek lisans tezi). Marmara Üniversitesi, İstanbul.