

Öğrenci Bilgi Seviyesinin Veri Madenciliği Yöntemleri ile Belirlenmesi ve Tahmin Edilmesi

Sezer Seven^{1*}, Tolga AYDIN² and Güler KARAMAN³

¹Atatürk Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Erzurum, Türkiye

²Atatürk Üniversitesi Mühendislik Fakültesi, Erzurum, Türkiye

³Ankara Hacı Bayram Veli Üniversitesi İktisadi ve İdari Bilimler Fakültesi, Ankara, Türkiye

*(sezerseven@atauni.edu.tr)

Özet – Öğretmenler açısından öğrencilerin herhangi bir konuda bilgi düzeylerinin bilinmesi oldukça önemlidir. Bu doğrultuda öğrencilerin bilgi düzeylerinin ölçülmesi veya belirlenmesine yönelik olarak çalışmalar literatürde mevcuttur. Yapılan çalışmalara genel olarak bakıldığında öğrenci bilgi düzeyinin belirlenebilmesi için nicel veya nitel araştırma yöntemlerinin kullanıldığı görülmüştür. Bu makalede gerçekleştirilen çalışmada ise öğrenci bilgi seviyesinin veri madenciliği yöntemleri kullanılarak tespit edilebilmesi amaçlanmıştır. Bilgi seviyelerinin oluşturulması işlemi için bölümleyici yöntemler arasında en çok kullanılan k-ortalamlar yöntemi kullanılmıştır. Çalışmadaki veri seti için, bilgi seviyelerinin optimum olarak 4 gruba ayrılması gerektiği görülmektedir. Bilgi seviyesinin tahmini için ise regresyon analizi tercih edilmiştir. Bu çalışma için sınıflandırma temelli olan lojistik regresyon modeli, sınıflandırma ve regresyon ağaçları modeli ve son olarak destek vektör regresyon modelleri kullanılmıştır. Sonuç olarak lojistik regresyon modeli %87, sınıflandırma ve regresyon ağaçları modeli %84.38 ve son olarak destek vektör regresyon modeli %82.66 oranında başarılı bir şekilde tahmin gerçekleştirmiştir.

Anahtar Kelimeler – Bilgi Seviyesi; Regresyon Modeli; Veri Madenciliği; K-Ortalamlar; Sınıflandırma

I. GİRİŞ

Öğretmenler açısından öğrencilerin herhangi bir konuda bilgi düzeylerinin bilinmesi oldukça önemlidir. Bu doğrultuda öğrencilerin bilgi düzeylerinin ölçülmesine veya belirlenmesine yönelik olarak çalışmalar gerçekleştirilmiştir. Bu çalışmalara genel olarak bakıldığında öğrenci bilgi düzeyinin belirlenebilmesi için nicel veya nitel araştırma yöntemleri kullanılmıştır. Örneğin; Gökdere ve Ayvaci (2004) yapmış oldukları sınıf öğretmenlerinin üstün yeteneklilik kavramı hakkındaki bilgi düzeylerini belirlemeyi amaçladığı çalışmada başarı testi ve mülakat yöntemi kullanılmıştır. Karamustafaoglu ve Yalnız'ın (2016) yapmış oldukları 8. sınıf öğrencilerinin genel kültür düzeylerinin belirlenmesini amaçladığı çalışmada ise nicel araştırma yöntemi olan betimsel alan taraması kullanılmıştır. Son olarak; İncekara ve Tuna'nın (2010) yapmış ortaöğretim seviyesinde öğrenim gören öğrencilerin çevre ile ilgili olgular hakkındaki bilgi seviyelerinin belirlenmesini amaçlayan çalışmada nicel araştırma yöntemi olan ölçek yöntemi kullanılmıştır.

Bu motivasyondan yola çıkılarak yapılan bu çalışmada öğrenci bilgi seviyesinin veri madenciliği yöntemleri kullanılarak tespit edilebilmesi amaçlanmıştır. Bu doğrultuda yapılan alanyazın taraması neticesinde benzer bir çalışmaya rastlanılmamıştır. Öncelikle bilgi seviyelerinin oluşturulması için benzer özellik gösteren öğrenci gruplarının oluşturulması fikri üzerinde çalışmalar gerçekleştirilmiştir. Öğrencileri bilgi seviyelerine göre ayırmak, benzer bilgi düzeyi özelliği gösteren öğrencilerin bir araya toplanarak gruplandırılması demektir. Bu durumdan ötürü grupların oluşturulma işlemi için bölümleyici yöntemlerin arasında en çok kullanılan k-ortalamlar yöntemi (Cebeci ve ark., 2015) kullanılmıştır. K-

ortalamlar yöntemi algoritmik yapısı gereği k yani küme sayısının dışarıdan girilmesini istemektedir. Bu çalışmada küme sayısı belli olmadığından ötürü dirsek yöntemi kullanılmıştır. Dirsek yöntemi ile optimum k sayısı bulunabilmektedir (Okumuş ve ark., 2017).

Öğrenciler bilgi düzeylerine göre gruplandırıldıktan sonra veri setine yeni dahil olacak öğrenciler için bilgi düzeylerinin tahmin edilmesi aşamasına geçilmiştir. Tahminleme için yaygın olarak kullanılan bir makine öğrenmesi modeli olan regresyon modeli tercih edilmiştir. Makine öğrenmesi algoritmalarında bir çok regresyon modeli kullanılmaktadır. Bu çalışmada ise üç farklı regresyon modeli kullanılmıştır. Bunlar; lojistik regresyon modeli, sınıflandırma ve regresyon ağaçları modeli ve son olarak destek vektör regresyon modelidir. Kullanılan regresyon modelleri sınıflandırma temelli olduğu ve yüksek tahmin doğruluğu gerçekleştirdiği için tercih edilmiştir (Awad ve Khanna, 2015; Işığışık, 2003; akt. Girginer ve Cankuş, 2008; Timofeev, 2004). Kullanılan üç regresyon modelinin başarı oranı birbirleri ile kıyaslanarak en yüksek başarı oranına sahip olan regresyon modeli belirlenmiştir.

K-Ortalamlar Yöntemi

K-ortalamlar yöntemi bölümleyici yöntemlerin arasında en çok kullanılan yöntemdir. Yöntemin isminden de anlaşılacağı üzere veri setini k adet kümeye bölümlemek için başlangıç olarak rastgele şekilde k adet küme merkezi seçilmektedir. Her bir veri noktasının küme merkezleri ile arasındaki uzaklıklar hesaplanmaktadır. Veri noktasının uzaklığı hangi kümeye yakın olursa ilgili kümeye dâhil edilmektedir. Veri noktalarının artık küme değiştirme durumu gözlemlenmiyorsa kümeleme işlemi sona erdirilir, gözlemleniyor ise bir sonraki

adımda işlem tekrarlanmaktadır (Cebeci ve ark., 2015). K-Ortalamalar yöntemi algoritmik yapısı gereği k sayısının dışarıdan girilmesini istemektedir. Küme sayısı belli olan veri setleri dışında belli olmayan veri setleri de mevcuttur. Dirsek yöntemi ile optimum k sayısı bulunabilmektedir. Yöntemde veri setindeki karesel hata oranları grafik eğrisi şeklinde değerlendirilmektedir. Eğride dirseğe benzemekte olan keskin bir kırılmanın yaşandığı nokta tespit edilmektedir. Bu nokta veri setinin optimum küme sayısını vermektedir (Okumuş ve ark., 2017).

Destek Vektör Regresyon Modeli

Destek Vektör Regresyon modeli, sınıflandırma problemlerinde kullanılan, doğrusal veya doğrusal olmayan bir fonksiyonun kullanılmasıyla sınıflandırma işlemini gerçekleştirmektedir (Özkan, 2020). Destek Vektör Regresyon modelinin, gerçek değer fonksiyonu tahmininde etkili bir araç olduğu kanıtlanmıştır. Ana avantajlarından biri, yüksek tahmin doğruluğu ile mükemmel genelleme yeteneğine sahip olmasıdır (Awad ve Khanna, 2015).

Lojistik Regresyon Modeli

Lojistik regresyon analizinin temel odağı, bireylerin hangi grubun üyesi olduğunu kestirmede bir regresyon denklemi oluşturmaktır (Çokluk, 2010). Lojistik regresyon analizinin iki amacı bulunmaktadır. Bu amaçlardan birincisi sınıflandırma, ikincisi ise bağımlı ve bağımsız değişkenler arasındaki ilişkileri araştırmaktır (Işığışık, 2003; akt. Girginer ve Cankuş, 2008).

Sınıflandırma ve Regresyon Ağaçları (CART) Modeli

Sınıflandırma ve Regresyon Ağaçları, bir ağaç gibi kökünden başlayarak baş aşağı büyürken gösterilmektedir. Bu algoritmada ikili dallanmalar söz konusudur yani bir düğümde seçme işlemi yapıldığında sadece iki dala ayrılabilir. Bir düğüme belirli bir kriter uygulanarak bölünme işlemi gerçekleştirilmektedir ve bu bölünmeler üzerinde seçme işlemi gerçekleştirilmektedir (Özkan, 2020). Sınıflandırma ve Regresyon Ağaçları, verileri sınıflandırmak için kullanılmaktadır. Bu algoritmayı kullanabilmek için sınıf sayısını önceden bilmemiz gerekmektedir (Timofeev, 2004).

II. MATERYAL VE YÖNTEM

Öğrenci bilgi seviyesinin belirlenmesi ve tahmin edilmesini amaçlayan bu çalışmada, mobil öğrenme ortamlarından elde edilen metriklerden yola çıkılarak öğrencilerin bilgi seviyelerinin veri madenciliği yöntemleri ile tahmin edilebileceği düşünülmüştür. Yapılan alanyazın taraması sonrasında 18 adet metrik elde edilmiştir. Bilgi seviyesi tespiti için metriklerin belirlenmesi aşamasından sonra bu metriklerin referans değerlerinin belirlenmesi aşamasına geçilmiştir. Bu aşamada nitel araştırma yöntemlerinden doküman inceleme yöntemi kullanılmıştır.

Doküman inceleme yöntemi, araştırılması hedeflenen olgular hakkında bilgi içeren yazılı materyallerin analizini kapsamaktadır. Nitel araştırmalarda kullanılması gereken önemli bilgi kaynaklarıdır. Araştırmacının gözlem ve görüşme yapmadan ihtiyacı olan veriyi elde edebilmesini sağlamaktadır. Veri toplama yöntemleriyle birlikte kullanıldığında verinin çeşitlendirilmesinde ve araştırmanın geçerliliğinin artırılmasında etkili olmaktadır. Eğitim ile ilgili yapılan araştırmalarda öğrenci kayıtları ve tutanakları, öğrenci ödevleri ve sınavları, öğretmen dosyaları doküman ve veri

kaynağı olarak kullanılabilir (Yıldırım & Şimşek, 1999).

Doküman inceleme yönteminden yola çıkılarak çalışmada kullanılan metriklere ait verilerin kişilerle herhangi bir görüşme sağlanmadan ya da anket uygulamadan elde edilebileceği görülmüştür. Böylelikle yapılan çalışmada, gözlem veya görüşme yapılmaksızın verilerin elde edilebilmesini sağlayabileceği ve eğitim araştırmalarında kullanılan önemli yöntemlerden biri olduğundan ötürü doküman inceleme yöntemi tercih edilmiştir.

Metriklerin oluşturulabileceği veri noktalarına yönelik olarak öğrencilerin mobil öğrenme ortamlarındaki faaliyetleri ile öğretmen tarafından öğrencilerle ilgili olarak düzenli bir şekilde tutulan kitap okuma sıklığı, çözülen soru sayısı, doğru veya yanlış sayısı gibi verileri barındıran dokümanlar incelenmiştir. Dokümanların incelenmesi sonucunda metriklerin taban ve tavan değerleri belirlenmiştir. Taban ve tavan değerler referans alınarak her bir öğrenci için simülasyon veriler oluşturulmuştur.

Sonuç olarak öğrencilerin önce k-ortalama yöntemine göre bilgi seviyesi ölçülerek, bilgi seviyesine göre gruplandırma işlemi gerçekleştirilmiştir. Sonrasında ise öğrencilerin bilgi seviyeleri sınıflandırma temelli üç regresyon modeli ile tahmin edilmeye çalışılmıştır. Yöntem kısmında yapılacak olan bu iki önemli işlem, veri bilimi başta olmak üzere diğer alanlarda da sıklıkla kullanılan Python programlama dili kullanılarak gerçekleştirilmiştir.

Araştırmanın Evren ve Örneklemi

Araştırmanın evrenini, 2021-2022 eğitim-öğretim yılı birinci döneminde, Erzurum iline bağlı tüm ilçelerdeki devlet okullarında ilköğretim kademesinde eğitim gören 7. Sınıf öğrencileri oluşturmaktadır. Bu örneklemin tercih edilme sebebi benzer özellik gösteren öğrenci gruplarının bölümleyici yöntemler ile gruplara daha sağlıklı bir şekilde ayrıştırılmasını sağlamaktır. Evrendeki öğrenci sayısı toplamda 12307 kişi olarak tespit edilmiştir. Araştırmanın örneklemini yaş aralığı 12 ila 13 arasında değişen 200'ü erkek, 172'si kız olmak üzere toplam 372 katılımcı oluşturmaktadır. Örneklem sayısının belirlenmesine yönelik olarak aşağıda belirtilen basit seçkisiz örnekleme formülü kullanılmıştır (Nacar, 2019). Formüle göre %95 güven aralığı ve 0,05 hata payına sahip olmak üzere, evren büyüklüğü 12307 olan araştırmalarda örneklem sayısının en az 372 olması gerekmektedir. Örneklemin hesaplanmasında kullanılan formül aşağıda belirtilmiştir (Karasar, 2014).

n: Örneklem büyüklüğü

N: Evren büyüklüğü

t: Belirli bir anlamlılık düzeyinde t tablosuna göre bulunan teorik değer

p: İncelenen olayın gerçekleşme olasılığı

q: İncelenen olayın gerçekleşmeme olasılığı

d: Hata payı

$$n = \frac{Nt^2pq}{d^2(N-1) + t^2pq}$$

$$= \frac{12307 \times 1,96^2 \times 0,5 \times 0,5}{0,05^2 \times (12307 - 1) + 1,96^2 \times 0,5 \times 0,5} = 372$$

Bilgi Seviyesi Metriklerinin Referans Değerlerinin Belirlenmesi

Aşağıdaki bahsi geçen metriklerin referans değerleri mobil öğrenme ortamlarından elde edilen veriler ve öğrencilerin performansları ile ilgili verilerin düzenli olarak tutulduğu dokümanlar yardımı ile belirlenmiştir. Referans değerleri hesaplanırken veriler haftalık olarak belirlenmiştir.

- **Favori ders:** Milli Eğitim Bakanlığı 7. Sınıf ders listesinden faydalanılarak zorunlu olan 16 ders, referans değeri olarak belirlenmiştir.
- **Okul not ortalaması:** 45 ila 100 arasında referans değeri belirlenmiştir.
- **Ders not ortalaması:** 15 ila 100 arasında referans değeri belirlenmiştir.
- **Oturum süresi:** 16 hafta boyunca mobil öğrenme ortamında geçirilen toplam süre olarak dakika cinsinden hesaplanmıştır. Her hafta için taban değer öğrencinin toplam etkileşim süresi (video, okuma ve soru materyalinde geçirilen toplam süre); tavan değer ise 840 dakika olarak belirlenmiştir.
- **Ders süresince atılan mesaj sayısı:** 16 hafta boyunca mobil öğrenme ortamında atılan toplam mesaj sayısı olarak hesaplanmıştır. Her hafta için taban değer 0, tavan değer ise 50 adet olarak belirlenmiştir.
- **Ders süresince izlenen video süresi:** 16 hafta boyunca mobil öğrenme ortamında izlenen toplam video süresi olarak dakika cinsinden hesaplanmıştır. Ortalama bir video 15 dakika olmak üzere her hafta için maksimum izlenebilecek video sayısı 4 olarak belirlenmiştir. Bundan ötürü her hafta için taban değer 0, tavan değer ise 60 dakika olarak belirlenmiştir.
- **Soru başına ortalama ayırdığı süre:** 16 hafta boyunca mobil öğrenme ortamında çözülen toplam soru sayısı ve sorulara ayrılan dakika cinsinden toplam süre hesaplanmıştır. Daha sonra ise toplam süre, toplam soru sayısına bölünmüştür.
- **Toplam etkileşim süresi:** 16 hafta boyunca mobil öğrenme ortamında video, okuma ve soru materyalinde geçirilen dakika cinsinden toplam süre olarak hesaplanmıştır.
- **Günlük çözülen ortalama soru sayısı:** 16 hafta boyunca mobil öğrenme ortamında çözülen toplam soru sayısının 112'ye bölünmesi ile hesaplanmıştır.
- **Tekrarlayan okumalar:** 16 hafta boyunca mobil öğrenme ortamında en çok okunan materyalin yineleme sayısı olarak hesaplanmıştır.
- **Okuma için ayırdığı ortalama süre:** 16 hafta boyunca mobil öğrenme ortamında okuma materyali için ayrılan dakika cinsinden toplam sürenin ortalaması alınarak hesaplanmıştır.
- **Konu başına tamamlanan okuma sayısı:** 16 hafta boyunca mobil öğrenme ortamında yapılan toplam okuma sayısının konu sayısına(16) bölünmesi ile hesaplanmıştır.
- **Konu başına tamamlanan video sayısı:** 16 hafta boyunca mobil öğrenme ortamında izlenen toplam video sayısının konu sayısına(16) bölünmesi ile hesaplanmıştır.
- **Konu başına tamamlanan soru sayısı:** 16 hafta boyunca mobil öğrenme ortamında çözülen toplam

soru sayısının konu sayısına (16) bölünmesi ile hesaplanmıştır.

- **Konu başına doğru sayısı:** 16 hafta boyunca mobil öğrenme ortamında çözülen sorulardan elden edilen toplam doğru sayısının konu sayısına (16) bölünmesi ile hesaplanmıştır.
- **Konu başına yapılan yorum sayısı:** 16 hafta boyunca mobil öğrenme ortamında yapılan toplam yorum sayısının konu sayısına (16) bölünmesi ile hesaplanmıştır.
- **Görev tamamlama zamanı:** 16 hafta boyunca mobil öğrenme ortamında tamamlanan toplam görev sayısının konu sayısına (16) bölünmesi ile hesaplanmıştır.

Bilgi Seviyelerinin Oluşturulması ve Tahmin Edilmesi

Bilgi seviyelerinin oluşturulması için öncelikle benzer özellik gösteren öğrenci grupları bölümleyici yöntemler arasında en çok kullanılan k-ortalama yöntemini kullanılarak, Python programlama dili ve Python programlama diline ait pandas ve sklearn.cluster kütüphanelerinden yararlanılmıştır. Yöntemin isminden de anlaşılacağı üzere veri setini k adet kümeye bölümlenmek için başlangıç olarak rastgele şekilde k adet küme merkezi seçilmektedir. Her bir veri noktasının küme merkezleri ile arasındaki uzaklıklar hesaplanmaktadır. Veri noktasının uzaklığı hangi kümeye yakın olursa ilgili kümeye dâhil edilmektedir. Veri noktalarının artık küme değiştirme durumu gözlemlenmiyorsa kümeleme işlemi sona erdirilir, gözlemleniyor ise bir sonraki adımda işlem tekrarlanmaktadır (Cebeci ve ark., 2015).

K-Ortalama yöntemini ilgili yoğun bir şekilde merak edilen konulardan birisi ise optimum k sayısının bulunması işlemidir. K-Ortalama yöntemini algoritmik yapısı gereği k sayısının dışarıdan girilmesini istemektedir. Küme sayısı belli olan veri setleri dışında belli olmayan veri setleri de mevcuttur. Bu durumda devreye dirsek(elbow) yöntemi girmektedir. Dirsek yöntemi ile optimum k sayısı bulunabilmektedir. Yöntemde veri setindeki karesel hata oranları grafik eğrisi şeklinde değerlendirilmektedir. Eğride dirseğe benzemekte olan keskin bir kırılmanın yaşandığı nokta tespit edilmektedir. Bu nokta veri setinin optimum küme sayısını vermektedir (Okumuş ve ark., 2017).

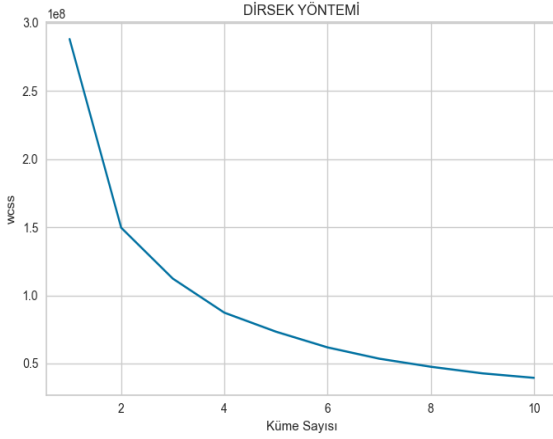
Bilgi seviyelerinin tahmin edilmesi işlemi için sınıflandırma temelli olan üç ayrı regresyon modeli kullanılmıştır. Bunlar; lojistik regresyon modeli, sınıflandırma ve regresyon ağaçları modeli ve son olarak destek vektör regresyon modelidir. Veriler, regresyon analizine tabi tutulmadan önce min-max normalizasyon işlemi uygulanmıştır. Regresyon analizleri Python programlama dili kullanılarak sklearn kütüphanesi yardımıyla gerçekleştirilmiştir. Modellerin başarı oranına R² skoruna bakılarak karar verilmiştir. Her bir modelin başarı oranı birbiri ile kıyaslanmıştır.

III. BULGULAR

Çalışmanın bu bölümünde öğrenci bilgi seviyesinin belirlenmesi ve tahmin edilmesine yönelik olarak bulgular sunulmuştur.

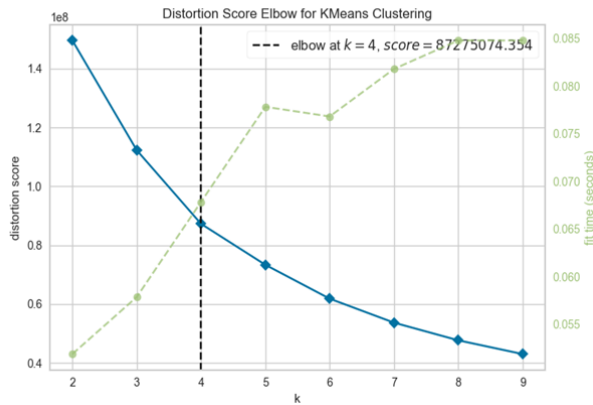
Öğrenci Bilgi Seviyelerinin Belirlenmesi

Yapılan bu çalışmada kullanılan veri setinin optimum küme sayısı Python programlama dili kütüphanelerinden matplotlib kütüphanesi vasıtasıyla ile dirsek yöntemi kullanılarak elde edilen grafik eğrisi sayesinde tespit edilmiştir. Aşağıda Şekil 1.1'de görüldüğü üzere optimum k sayısı 4 olarak gösterilmektedir.



Şekil. 1.1. Dirsek Yöntemi ile Optimum K Değeri Grafliği

Grafik eğrisi ile optimum k sayısının tespit edilememesi durumunda yine Python programlama dili kütüphanelerinden klibowvisualizer kütüphanesi vasıtasıyla ile dirsek yönteminin görselleştirilmiş versiyonu kullanılabilir. Aşağıda Şekil 1.2'de görüldüğü üzere dirsek yönteminin görselleştirilmiş versiyonu kullanılarak optimum k sayısı tekrardan 4 olarak tespit edilmiştir.



Şekil. 1.2. Görselleştirilmiş Optimum K Değeri Grafliği

Öğrenci Bilgi Seviyelerinin Tahmin Edilmesi

Bilgi seviyelerinin tahmin edilebilmesi için üç ayrı regresyon modeli olan "Lojistik regresyon", "Destek Vektör Regresyon" ve "Sınıflandırma ve Regresyon Ağaçları" kullanılmıştır. Yapılan regresyon analizleri sonucunda modellerin başarı puanı olarak R^2 skorlarına bakılmıştır. Tablo 1'de regresyon modellerinin başarı puanları sıralanmaktadır.

Table 1. Regresyon Modelleri Başarı Puanı Sıralaması

S.N.	Model Adı	Başarı Puanı (R^2)
1	Lojistik Regresyon Modeli	0.86992
2	Sınıflandırma ve Regresyon Ağaçları	0.84379
3	Destek Vektör Regresyon Modeli	0.82653

IV. SONUÇ

Çalışmanın bu bölümünde bulgulara ait sonuçlar tartışılmaktadır. Ayrıca gelecek çalışmalara ışık tutacak öneriler sunulmaktadır.

Elde edilen bulgulara bakıldığında çalışmada kullanılan veri setindeki öğrencilerin bilgi seviyelerinin optimum olarak 4 gruba ayrılması gerektiği Şekil 1.1. ve Şekil 1.2.'de görülmektedir. Gelecekte başka veri setleriyle yapılacak olan diğer çalışmalarda da optimum bilgi seviyesi sayısının dirsek yöntemi kullanılarak belirlenebileceği görülmüştür.

Öte yandan bilgi seviyesi tahmini ile ilgili olarak lojistik regresyon modeli yaklaşık %87, sınıflandırma ve regresyon ağaçları modeli yaklaşık %84.38 ve son olarak destek vektör regresyon modeli de yaklaşık %82.66 oranında başarılı bir şekilde öğrenci bilgi seviyelerini tahmin işlemi gerçekleştirmiştir. Böylelikle veri setine yeni dahil olan bir öğrencinin bilgi düzeyi lojistik regresyon modeli kullanılarak yüksek doğruluk oranı ile tahmin edilebilecektir.

Gelecekte yapılacak olan çalışmalarda elde edilecek veri çeşitliliği ile grupların daha kesin bir şekilde tahmin edilmesi sağlanabilir. Ayrıca veriler kısa süreli periyotlarla güncellenerek grupların dağılımı anlık olarak güncellenebilir.

KAYNAKLAR

- [1] Çokluk, Ö. (2010). Lojistik regresyon analizi: Kavram ve uygulama. Kuram ve uygulamada eğitim bilimleri, 10(3), 1357-1407.
- [2] Timofeev, R. (2004). Classification and regression trees (CART) theory and applications. Humboldt University, Berlin, 54.
- [3] Gökdere, M., & Ayvaci, H. Ş. (2004). Sınıf öğretmenlerinin üstün yetenekli çocuklar ve özellikleri ile ilgili bilgi seviyelerinin belirlenmesi. Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi, (18), 17-26.
- [4] Karamustafaoglu, S., & Yalnız, S. (2016). 8. Sınıf öğrencilerinin genel kültür seviyelerinin belirlenmesi.
- [5] İncekara, S., & Tuna, F. (2010). Ortaöğretim Öğrencilerinin Çevresel Konularla İlgili Bilgi Düzeylerinin Ölçülmesi: Çankırı İli Örneği. Marmara Coğrafya Dergisi, (22), 168-182.
- [6] Timofeev, R. (2004). Classification and regression trees (CART) theory and applications. Humboldt University, Berlin, 54.
- [7] Özkan, Y. (2020). Veri madenciliği yöntemleri. Papatya Yayıncılık Eğitim.
- [8] Girginer, N., & Cankuş, B. (2008). Tramvay yolcu memnuniyetinin lojistik regresyon analiziyle ölçülmesi: Etram örneği. Yönetim ve Ekonomi Dergisi, 15(1), 181-193.
- [9] Işığçok, E. (2003). Bebeklerin doğum ağırlıklarını ve boylarını etkileyen faktörlerin lojistik regresyon analizi ile araştırılması. IV. Ulusal Ekonometri ve İstatistik Sempozyumu, 3.
- [10] Awad, M., & Khanna, R. (2015). Support vector regression. In Efficient learning machines (pp. 67-80). Apress, Berkeley, CA.
- [11] Cebeci, Z., Yıldız, F., & Kayaalp, G. T. (2015). K-ortalamlar kümelemesinde optimum K değeri seçilmesi. 2. Ulusal Yönetim Bilişim Sistemleri Kongresi. Erzurum, 8-10.
- [12] Pınar, M., Okumuş, O., Turgut, U. O., Kalıpsız, O., & Aktaş, M. S. (2017). Büyük veri içeren öneri sistemleri için hiperparametre optimizasyonu. Ulusal Yazılım Mühendisliği Sempozyumu, 22-272.
- [13] Yıldırım, A., & Simsek, H. (1999). Sosyal Bilimlerde Nitel Araştırma Yöntemleri (12. baskı: 1999-2021).
- [14] Nacar, A. (2019). Öğretmenlerin Eğitime İnanma Düzeyleri ile İşe Bağlılık Düzeyleri Arasındaki İlişki: Kahramanmaraş İli Örneği. (Yüksek Lisans Tezi). Kahramanmaraş: Sütçü İmam Üniversitesi Sosyal Bilimler Enstitüsü.
- [15] Karasar, N. (2014). Bilimsel Araştırma Yöntemi: Kavramlar, İlkeler, Teknikler. 26. bs. Ankara: Nobel Yayın Dağıtım