

Çocuklar İçin Otizm Spektrum Bozukluğunun Veri Madenciliği Yöntemleri ile Analizi*

Analysis of Autism Spectrum Disorder in Children with Data Mining Methods

Sümeyye Çelik¹ , Melike Şişeci Çeşmeli² 



*Bu çalışma "Çocuklar İçin Otizm Spektrum Bozukluğunun Veri Madenciliği Yöntemleri ile Analizi" başlıklı yüksek lisans tezinden üretilmiştir.

¹Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi, İşletme Fakültesi, Yönetim Bilişim Sistemleri Bölümü, Adana, Türkiye

²Burdur Mehmet Akif Ersoy Üniversitesi, Bucak Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü, Burdur, Türkiye

ORCID: S.Ç.0000-0003-3621-8653;
M.Ş.Ç. 0000-0001-9541-2590

Corresponding author:

Sümeyye ÇELİK
Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi, İşletme Fakültesi, Yönetim Bilişim Sistemleri Bölümü, Adana, Türkiye
E-mail address: sumeyyecelik@atu.edu.tr

Submitted: 17.02.2021

Revision Requested: 06.04.2021

Last Revision Received: 08.04.2021

Accepted: 08.04.2021

Published Online: 29.06.2021

Citation: Çelik, S. ve Şişeci Çeşmeli, M. (2021). Çocuklar için otizm spektrum bozukluğunun veri madenciliği yöntemleri ile analizi. *Acta Infologica*, 5(1), 167-186. <https://doi.org/10.26650/acin.882187>

ÖZ

Veri madenciliği teknikleri, veriler arasında gizli kalmış olan örüntüleri ortaya çıkarmayı amaçlamaktadır. Bu kapsamda, tıp gibi birçok alanda yaygın bir biçimde kullanılmaktadır. Teşhis ve tedavisi oldukça zor ve uzun bir süreçten oluşan otizm spektrum bozukluğu doğuştan gelen ya da yaşamın ilk yıllarında ortaya çıkan karmaşık bir nöro-gelişimsel bozukluktur. Bu çalışmada 292 çocukta toplanan gerçek ve güncel otizm spektrum bozukluğu verileri kullanılmıştır. Veri seti 20 girdi özneliği ve 1 çıktı özneliğine sahiptir. Çıktı özneliği otizmin bulunup bulunmadığını ifade etmektedir. Çalışma da öncelikle veri seti üzerinde eksik verilerin tamamlanması, kategorik verilerin sayısallaştırılması, normalizasyon gibi veri ön işleme aşamaları gerçekleştirilmiştir. Devamında ise öznelikler yapay sinir ağları ve dilsel kuvvetli sinir-bulanık sınıflayıcı ile sınıflandırılmış, k-means ve x-means ile kümelendirilmiştir. Her bir yöntemin sonuçları değerlendirilmiş ve performanslar karşılaştırılmıştır.

Anahtar kelimeler: Tıp Alanında Veri Madenciliği, Hastalık Teşhisi, Otizm Spektrum Bozukluğu, Sınıflandırma, Kümeleme

ABSTRACT

Data mining techniques aim to reveal hidden patterns in data. They are widely used in many fields, such as medicine. Autism spectrum disorder, whose diagnosis and treatment are difficult and lengthy, is a complex neurodevelopmental disorder that is congenital or occurs in the first years of life. Actual and current autism spectrum disorder data collected from 292 children were used in this study. The data set has 20 input attributes and 1 output attribute. The output attribute expresses whether autism is present or not. In the study, data pre-processing stages, such as completing missing data on the data set, digitizing categorical data, and normalization, were first carried out. Subsequently, the features were classified by artificial neural networks and linguistic strength neuro-fuzzy classifier and clustered with k-means and x-means. The results of each method were evaluated and the performances were compared.

Keywords: Data Mining in Medicine, Diagnosis of Disease, Autism Spectrum Disorder, Classification, Clustering

1. GİRİŞ

Veri madenciliği, bilgiyi keşfetme yolculuğudur (Hong ve Wu, 2011). Veri madenciliği ile var olan ama daha keşfedilmemiş örüntüler aranmaktadır. Yani veri madenciliği; ham halde iken sadece yığın şeklinde duran ve bir anlam ifade etmeyen veriler arasından, amaca uygun olarak anlamlı ifadeler oluşturulmasına olanak tanıyan yöntemler bütünüdür. Veri madenciliğinin en temel amacına bakıldığında zaman cevher madenciliği ile aynı olduğu görülmektedir (Diler, 2016).

Teknolojik gelişmeler her alanda olduğu gibi tıp alanında da oldukça kullanılmaktadır (Moskov, 2016). Günümüzde gelişen bu teknoloji ile birlikte hastanelerde kullanılan cihazlar ve uygulanan tedavi yöntemleri de geliştirilmektedir. Donanımsal ve yazılımsal olarak sürekli gelişen bu yöntemler sayesinde teşhis ve tedavilerin daha profesyonel şekilde uygulanmasına olanak tanınmaktadır (Tsumoto vd., 2017).

Tıp alanında söz konusu insan sağlığı olduğu için Ar-Ge çalışmalarının olabildiğince fazla olması ve desteklenmesi önemlidir (Reddy vd., 2012). Birçok tedavide erken teşhis büyük bir öneme sahiptir. Tedaviye geç kalındığında zaman hastalık daha fazla ilerleyebilmekte ve tedavisi daha zor olabilmektedir. Hatta bazı durumlarda geri dönüşümü mümkün olmayacak kayıplara sebep olabilmektedir. Bu gibi sebeplerden dolayı teşhis ve tedavi noktasında doktora yardımcı olması için kullanılan veri madenciliği yöntemlerine literatürde sıkça rastlanmaktadır (Akdemir, 2016).

Tıp alanında yapılacak olan teknolojik çalışmalar birçok farklı disiplinin bir araya gelmesi ile oluşturulmaktadır. Çünkü hastalık ve teşhis noktasında uzman olan bir ya da birkaç hekim olmadan hastadan gerekli bilgiler alınamayacak ya da muayene sonucunda elde edilen bulgular değerlendirilemeyecektir. Hastalık ile ilgili olarak elde edilen bulguların değerlendirmesini alanında uzman olmayan kişilerin yapması mümkün değildir. Her bir hastalık ayrı uzmanlık gerektirir ve hastalığın kişide olup olmadığının tespit edilmesi için spesifik detayların uzman tarafından biliniyor olması gereklidir. Aksi takdirde teşhis yanlış konulabilmekte ve tedavi aksayabilmektedir (Budak ve Bozkurt, 2013). Bu durum hastanın daha ciddi sorunlar yaşamasına sebep olabilmektedir. Sonuç olarak teşhis konusunda bilişim teknolojileri kullanılarak yapılacak bir çalışmada, bilgilerin alanında uzman kişilerden alınması gereklidir. Kullanılan parametreler doğruluğu uzmanlar tarafından kabul görmüş hali ile toplanmış olmalıdır ve analizler daha sonra gerçekleştirilmelidir. Bu analizleri gerçekleştirmek için de bilişim sistemlerini kullanabilen, uzmanından aldığı verileri işleyerek anlamlı bilgilere dönüştürebilecek kişilere ihtiyaç duyulmaktadır. Sonuç olarak başarılı bir çalışma yürütmek için farklı disiplinlerin bir arada çalışması oldukça önemlidir (Nilashi, 2017).

Otizm, Türkiye de ve dünyada oldukça fazla görülen bir rahatsızlıktır (Kılınç ve Söğüt, 2018). Otizm spektrum bozukluğu, doğuştan olabilen ya da doğumdan sonra bir kaç yıl içerisinde ortaya çıkabilen karmaşık bir nöro gelişimsel bozukluktur (Torun ve İşeri, 2018). Özellikle çocukluk döneminde görülmektedir (Uğur vd., 2018). Beynin yapısını ya da işleyişini etkileyen bazı sinir sistemi sorunlarından kaynaklanmaktadır (Torun ve İşeri, 2018). Otizm spektrum bozukluğu gözlemlenen hastalarda sosyal ve bilişsel olarak gelişmelerin normalden daha yavaş olduğu görülmektedir (Uğur ve Göker, 2018). Örneğin çevrelerinden gelen uyarılara karşı tepkileri yeterli düzeyde değildir (Gülsöz ve Çıkkılı, 2018). Birçok hastalıkta olduğu gibi otizmde de erken tedavi ve eğitim çok önemlidir. Bunun içinde erken teşhis önem kazanmaktadır. Çünkü çocukların hayattan uzak kalmadan, topluma kazandırılmaları gerekmektedir (Akman, 2018). Literatürde otizm spektrum bozukluğu konusunda birçok çalışma yer almaktadır. Bu çalışmaya da Thabtah (2017) tarafından gerçekleştirilen bilimsel çalışma yön vermiştir. Yazar çalışmasında otizm bozukluğu ile ilgili veri madenciliği çalışmalarını değerlendirmiştir. Yapılan araştırma ile son dönemde otizm verileri ile gerçekleştirilen veri madenciliği çalışmalarında verilerin güncelliğini yitirdiği ve yeni ölçeklendirmelerle toplanan verilere ihtiyaç duyuldu vurgusu yapılmıştır (Thabtah, 2017). Devamında ise Thabtah ve arkadaşları (2018) tarafından gerçekleştirilen bilimsel çalışma ile güncel olarak oluşturulan ölçekten faydalanılarak, ihtiyaç duyulan yeni veriler toplanmış ve sonuçlar yayınlanmıştır. Tarama testi ASDTests isimli bir mobil uygulama ortamında gerçekleştirilmiştir. Uygulamanın içinde yaş gruplarına göre ayrı kategoriler mevcuttur. Her kategoride on soru bulunmaktadır ve her biri kullanıcıların uygun cevabı dikkatlice seçmelerini sağlamak için bir resimle ilişkilendirilmiştir. Katılımcılara, bilgilerinin isimsiz tutulacağı ve sadece araştırma amacıyla paylaşılacağı bildirilmiştir. Değerlendirme tamamlanmadan önce, katılımcılardan araştırmanın amacını, gizlilik politikasını ve verilerin kullanımını açıklayan bir bilgilendirme

yapılmıştır. Kullanılan veri setinde girdi olarak kullanılan öznelikler arasında otizm spektrum bozukluğu'nu ölçümleyen on sorunun yanında demografik özellikleri içeren bilgilerde mevcuttur. Çıktı parametresi ise teste giren bireyin otizm spektrum bozukluğu olup olmadığını belirtmek için kullanılan evet ya da hayır anlamı taşıyan sınıf değişkenidir. (Thabtah vd., 2018).

Otizm spektrum bozukluğu için Thabtah ve arkadaşların oluşturduğu güncel verilerin çocuklar için olan alt kümesi kullanılarak yapılan çalışmalar incelendiği zaman logistic regression, naive bayes gibi yöntemlerin kullanılması ile gerçekleştirilen sınıflandırma işlemleri olmakla birlikte yapay sinir ağları ve dilsel kuvvetli sinir-bulanık sınıflayıcı ile yapılan bir sınıflandırmaya ve herhangi bir yöntem ile kümelemeye rastlanmamıştır. Bu sebepten dolayı bu çalışmada öncelikle eksik verilerin tamamlanması, normalizasyon işleminin gerçekleştirilmesi gibi veri ön işleme yöntemleri uygulanmış devamında ise bu konuda literatürde eksikliği görülen kümeleme işleminin gerçekleştirilmesi ve sınıflandırma işlemi için denenmediği fark edilen iki yöntemin veri seti üzerinde uygulanması sağlanmıştır. Literatüre katkısı olması açısından çalışmada denenilen yöntemlerin tahminliliğinin doğruluğu Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü gibi birçok parametre ile verilmiş ve yöntemlerin başarı oranları karşılaştırılmıştır. Kümeleme işlemi için k-means ve x-means yöntemleri, sınıflandırma işlemi için ise yapay sinir ağları ve dilsel kuvvetli sinir-bulanık sınıflayıcı yöntemleri kullanılmıştır. Sonuçlara bakıldığı zaman tahminleme doğruluğu bakımından otizm spektrum bozukluğu'nun çocuklar için olan verilerinde, sınıflandırma yöntemlerinin kümeleme yöntemlerinden daha başarılı olduğu görülmektedir. Özellikle dilsel kuvvetli sinir-bulanık sınıflayıcı yönteminin tüm verileri doğru sınıflandırarak literatürdeki birçok yöntemden daha yüksek başarı oranına sahip olduğu için tercih edilebilecek yöntemler arasında en iyi yöntemlerden bir tanesi olabileceği sonucuna varılmaktadır.

2. KAVRAMSAL ÇERÇEVE

2.1. Veri Madenciliği

Veri madenciliğinin literatürde birçok farklı tanımının olmasının yanında, en temel anlamına bakıldığında elde edilen verilerin veri madenciliği yöntemleri ile analiz edilerek anlamlandırılmaya çalışıldığı süreçtir. Veri madenciliği adından da anlaşıldığı üzere bir madenciliktir. Madencilik işinin çeşitli yöntemler yardımı ile verilere uygulanmasını sağlamaktadır (Akdemir, 2016). Böylece veriler üzerine uygulanan madencilik ile bilgi keşifleri gerçekleştirilmektedir. Yani var olan bilgiler, toplu halde duran ve ayrıştırılmamış olan veri yığınları içerisinde bulunarak keşfedilmektedir (Hong ve Wu, 2011).

2.2. Veri Madenciliğinin Tıp Alanında Kullanımı

Tıp alanında hastalık teşhisi yapılırken dikkate alınması gereken birçok kıstas vardır. Hastalığın ciddi olabileceği ve alınacak yanlış bir karar ile çok daha büyük problemlere yol açılabileceği göz önünde bulundurulduğunda teşhis konulurken hızlı ve doğru karar verilmesi oldukça önemlidir (Çelik, 2020). Bu noktada veri madenciliği yöntemlerinden faydalanmak gereklidir (Reddy vd., 2012). Tıp alanında, veri madenciliği hastalıklı dokuların tespit edilmesi ve tanı konulmasında ya da tedavi edilmesinde destek olacak sistemlerin tasarlanmasında kullanılabilir (Brossette vd., 1998).

2.3. Otizm Spektrum Bozukluğu

Otizm spektrum bozukluğu (OSB) dünya çapında oldukça hızlı büyüyen gelişimsel sakatlık tanısından biridir (Thabtah, 2017). OSB, doğuştan gelen ya da yaşamın ilk yıllarında ortaya çıkan karmaşık bir nöro-gelişimsel bozukluktur (Thabtah, 2018). Sosyal ilişkilerde, iletişim kurmada, öğrenmede ve konuşma yeteneklerinde sıkıntılar yaşanmasına sebep olmaktadır (De Campos Souza ve Guimaraes, 2018). OSB'de ilgi duyulan şeyler sınırlıdır ve bunun yanında kendini yineleyen davranışlarda ve hareketlerde bulunmaktadır (Akyol ve Karaci, 2018).

OSB'nin teşhis ve tedavisi oldukça uzundur. Aynı zamanda teşhisin erken konulması ve tedaviye başlanması çok önemlidir (Sabuncuoğlu vd., 2015). Bu hastalığın erken fark edilmesi ileride yaşanacak sağlık sorunlarını önemli ölçüde azaltabilmekte ve çocuğun genel zihinsel sağlığını iyileştirici yöntemlerin uygulanmasına imkân tanımaktadır (van den Bekerom, 2017). Teşhis için en güncel olarak Amerikan Psikiyatri Birliği'nin yayımladığı kılavuz (DSM-V) kullanılır (Olçay Gül ve Tekin İftar, 2012).

2.4. Otizm Spektrum Bozukluğu Konusunda Çocuklardan Toplanan Veri Seti ile Yapılan Çalışmalar

Literatür incelendiği zaman OSB konusunda veri madenciliği veya yapay zekâ tekniklerini içeren, OSB veri setinin farklı altkümelerini kullanan çok fazla çalışma olduğu görülmektedir. Bu sebepten dolayı literatürü sınırlandırabilmek için OSB veri setinin çocuklar için olan alt kümesini kullanan çalışmalara yer verilmiştir. Bu çalışmaların temelini topladığı verileri paylaşarak Thabtah (2018) oluşturmuştur.

Thabtah (2018) tarafından, detaylı olarak literatür değerlendirilmiş, daha önceki çalışmalara yer yerilmiştir. Verilerin nasıl toplandığı anlatılmış ve uygulama aracı olarak kullanılan ASDTests tanıtılmıştır. Veriler arasındaki ilişki incelenerek grafikler oluşturulmuştur. ASD için oluşturulan veri seti için üç alt küme olarak belirlenen çocuk, ergen, yetişkin grupları için Logistic Regression ve Naïve Bayes algoritmaları ile analizler gerçekleştirilmiştir. Elde sonuçlara göre çocuklar için elde edilen doğruluk Logistic Regression için % 97.94, Naïve Bayes için ise % 92.80 olmuştur.

Akyol ve Karaci (2018), Thabtah'ın çalışması ile oluşturulan güncel veri setinin çocuklar için olan alt kümesini kullanarak öznitelik seçimi gerçekleştirilmiş ve ardından sınıflandırılma yapılmıştır. Çalışmada öncelikle ön işleme aşamaları gerçekleştirilmiş ve boş değerler içeren örnekler veri setinden atılmıştır. Böylece toplamda 292 örnekten oluşan veri seti 247 örneğe indirilmiştir. Veri setindeki kategorik veri türüne sahip öznitelikler sayısal değerlere çevrilmiştir. Bu ön işleme aşamasından sonra sınıflandırma gerçekleştirilmiştir. Sınıflandırma için Lojistik Regresyon ve Bulanık Kurallı Lojistik Regresyon kombinasyonu kullanılmıştır. Kullanılan yöntemlerin başarı oranına bakıldığı zaman Lojistik Regresyon %92.00, Bulanık Kurallı Lojistik Regresyon kombinasyonu %97.33 oranında doğruluğa sahiptir. Sonuç olarak bu çalışma için lojistik regresyonun bulanık kurallar ile birlikte çalıştırıldığında daha doğru bir tahminleme yaptığı gösterilmektedir (Akyol ve Karaci, 2018).

De Campos Souza ve Guimaraes (2018), Thabtah'ın çalışması ile oluşturulan güncel veri setinin çocuklar için olan alt kümesini kullanarak bir sınıflandırma gerçekleştirmiştir. Sınıflandırma işlemi gerçekleştirilmeden önce veriler ön işleme aşamasından geçirilmiş ve kategorik olan veri türleri sayısal ifadelerle dönüştürülmüştür. Yöntem olarak Bulanık Sinir Ağı Mimarisi kullanılmıştır ve bulanık kümelerin sayısının performansa etkisi değerlendirilmiştir. Modelin başarı oranını test etmek için veriler %70 eğitim %30 test verisi olarak ayrılmıştır. Sonuçlara bakıldığı zaman en düşük performansın bulanık küme sayısının 3 olarak kullanıldığı test verisinde %59 olarak, en yüksek performansın da bulanık küme sayısının 9 olarak kullanıldığı test ve eğitim setinde %100 olarak görülmüştür (De Campos Souza ve Guimaraes, 2018).

Hutchinson ve arkadaşları (2019), Thabtah'ın çalışması ile oluşturulan güncel veri setinin çocuklar için olan alt kümesini kullanarak sınıflandırma yapmışlardır. Sınıflandırmayı gerçekleştirmek için yöntem olarak Weka hazır paket programı kullanarak J48 Decision Tree ve Naive Bayes algoritmasını kullanmışlardır. İki yöntem ile elde edilen sonuçlar accuracy, sensitivity ve specificity değerleri olarak verilmiş ayrıca iki yöntemin başarıları karşılaştırılmıştır. Çalışmanın sonuçlarına bakıldığı zaman; J48 Decision Tree için accuracy değerinin %92.8, sensitivity değerinin 0.901 ve specificity değerinin 0.934 olduğu Naive Bayes için ise accuracy değerinin %96.2, sensitivity değerinin 0.972 ve specificity değerinin 0.954 olduğu görülmektedir. Bu sonuçlar kapsamında kullanılan veri setini Naive Bayes algoritmasının J48 Decision Tree algoritmasından daha doğru sınıflandırdığı görülmektedir (Hutchinson vd, 2019).

3. YÖNTEM

3.1. Sınıflandırma Yöntemi

3.1.1. Yapay sinir ağları

Yapay sinir ağları (YSA), insan beyninin çalışma şekli ilham alınarak oluşturulmuş bir algoritmadır (Deb vd., 2018). Temelinde nörofizyologların ve psikologların sinir ağlarının nasıl çalıştığı konusunda ortaya koyduğu biyolojik bulgular kullanılmaktadır (Küçükkoçaoğlu vd., 2005). Bu biyolojik bulgular yapısal ve işlevsel olarak sistemleştirilerek matematiksel bir model oluşturulmaya çalışılmıştır. Oluşturulan bu modele sinir ağları modeli denilmiştir (Akkaya vd., 2009). Diğer bir

ifade ile YSA insan beyninin çalışma sistemini taklit ederek (Mehmet ve Akpınar, 2018), bilgisayarlı sistemlere karar verdirme becerisini kazandırmayı amaçlamaktadır (Kalantar vd., 2018). YSA, İnsan beyninin ezberleme, öğrenmeye, anlam çıkarma, bilgiler arasındaki ilişkileri ortaya koyma gibi birçok işlemini yapabilme amacıyla tasarlanmıştır. (Özsoy ve Fırat, 2004) (Akyılmaz ve Ayan, 2006). YSA ortadaki durumdan bir öğrenme gerçekleştirebildiği için yeni durumlara adapte olabileme yeteneğine sahiptir.

3.1.2. Dilsel Kuvvetli Sinir-Bulanık Sınıflayıcı

Dilsel kuvvetli sinir-bulanık sınıflayıcı (DKSBS), verileri sınıflandırmak için kullanılan yöntemlerden bir tanesidir. En temel özelliği sınıflandırma yapmadan önce özniteliklerin önem derecelerinin belirlenmesidir (Pençe vd, 2013). Özniteliklerin önem dereceleri belirlenirken bulanık çıkarsama yapılmaktadır. Böylece sınıflandırma sırasında, önem derecesi yüksek olan öznitelikler seçilirken, önem derecesi düşük olan öznitelikler devre dışı bırakılmış olmaktadır (Çetişli, 2006).

Klasik mantıkta ile bulanık mantığın en temel farkı; klasik mantıkta sınıflandırmaların kesin olması, bulanık mantıkta ise bir elemanın birden fazla kümeye ait olabilmesi durumudur. Bulanık mantıktaki elemanlar kümelere belirli oranlarda üye olabilmektedir. Yani elemanı olma durumunu 1, olmama durumunu 0 ile temsil edersek 0.3, 0.5 derecelerde elemanı olma durumu ortaya çıkabilmektedir. Bunun yanı sıra elemanlar tek bir kümede olmak zorunda da olmadığı için bir eleman 0.3 oranında bir kümeye, 0.5 oranında başka kümeye dahil olabilmektedir. Bu durum klasik kümelerdeki siyah beyaz şeklindeki keskin ayrımın ortadan kaldırılarak gri alanların meydana gelmesine olanak tanımaktadır. Böylece bulanık yaklaşım ile daha esnek davranılarak kısmi üyelikler oluşturulabilmektedir (Tiryaki ve Kazan, 2007).

Bulanık çıkarsama ile özniteliklerin önem dereceleri belirlendikten sonra sınıflandırma için YSA kullanılmaktadır. YSA eğitilirken, bulanık çıkarsama ile seçilen öznitelikler kullanıldığı için başarı oranı oldukça yüksektir (Çelik vd., 2018). Aynı zamanda kümeleri birbirinden ayırtırmayan öznitelikler kullanılmadığı için bellekten de tasarruf sağlanmaktadır. Bu durum hızın artmasının yanında, maliyetlerin de azalmasına yardımcı olmaktadır (Çetişli, 2006).

Sınıflandırma yapılırken bazı öznitelikler çıkarıldığında sınıflandırma başarısı düşmüyor ise o özniteliklerin sınıflandırmada kullanılmasına gerek yoktur (Huang ve Chow, 2005). Çünkü gerçek hayat problemlerinde de çözüme gidilirken oluşturulan alternatiflerin etki düzeyi farklı olabilmektedir. Önemli olan problemi çözüme götüreceği olan daha hızlı ve daha doğru alternatifleri bulmaktır (Liu vd., 2005). Bu açıdan bakıldığında zaman dilsel kuvvetli sinir-bulanık sınıflayıcıda da gerçek hayat problemlerinin çözümüne yakın bir çözüm yolu izlenmektedir. Veri madenciliği açısından izlenen yol ise iki aşamada gerçekleştirilmektedir. Öncelikle ön işleme aşaması olarak sayılan özniteliklerin önem derecesinin belirlenmesi işlemi gerçekleştirilmektedir. Daha sonra belirlenen bu önem düzeylerine göre yapay sinir ağı eğitilerek sınıflandırma işlemi gerçekleştirilmektedir (Çelik vd., 2018). Böylece dilsel kuvvetli sinir-bulanık sınıflayıcı'yı diğer sınıflandırma yöntemlerinden ayıran en önemli özelliği bulanık çıkarsama ile öznitelik tespiti gerçekleştirmesidir (Fırat, 2008).

Dilsel kuvvetli sinir-bulanık sınıflayıcı, özniteliklerin önem düzeyini hesaplarken bulanık kuralları kullanmaktadır. Öznitelikler, bulanık kurallar ile aşamalı olarak belirlenmektedir (Çağlar vd., 2010). Bulanıklık, sistemi evet hayır gibi kesin cevapların kısıtlarından kurtarmaktadır (Koçyiğit ve Korürek, 2010). Yani bir özneliğin sonuç üzerinde etkisinin olup olmadığı var ya da yok şeklinde kesin olarak nitelendirilmek yerine, “az var”, “var” yada “çok var” gibi aşamalı olarak nitelendirilmektedir. Böylece bulanık çıkarsamaların insan çıkarsamalarına benzerliği artırılmış olmaktadır (Cetişli ve Kalkan, 2011).

3.2. Kümeleme Yöntemi

3.2.1. K-Means

Kümeleme yöntemi, bir dizi elemanı benzer gruplarda toplama işlemidir. Kümeleme yöntemi veri setlerini gruplara ayırmak ve verilerdeki önemli ve gizli kalmış bilgileri açığa çıkarmak için en çok kullanılan yöntemlerden bir tanesidir (Çınaroğlu ve Bulut 2018). Kümeleme yönteminin, diğer bir veri madenciliği yöntemi olan sınıflandırmadan farkı şudur; sınıflandırmada sınıflar önceden bellidir ama kümelemede sınıflar önceden belli değildir (Doğan vd., 2018). K-means en eski kümeleme algoritmalarından bir tanesidir (MacQueen, 1967) ve hiyerarşik olmayan kümeleme yöntemleri arasında yer almaktadır

(Yavuz vd., 2011). 1967 yılında J.B. MacQueen tarafından geliştirilen algoritma (Işık ve Çamurcu, 2007), yıllardır birçok alanda kullanılmaya devam etmektedir (Bradley vd., 1998). K-means'ın tercih edilmesinin en temel sebebi yapısının sadeliği yanında model başarı oranının oldukça yüksek olmasıdır. K-means, kümeleri eşit büyüklükte ve homojen olarak oluşturma eğilimindedir (Işık ve Çamurcu, 2011) ve her elemanı bir kümeye atamaktadır. Böylece her eleman bir kümeye ait olmakta ve kümeler birbirinden kesin olarak ayrılmaktadır. Kümelemenin kullandığı öğrenme tekniği gözetimsiz öğrenmedir (Sariman, 2011). Yani kümeleme yapılırken önceden belirlenmiş kıstaslar yoktur. Kümeler, kümeyi temsil eden merkez noktaların belirlenmesi fikrine dayalı olarak oluşturulmaktadır (Güner vd., 2018). Kaç kümenin olacağı k parametresi ile ifade edilmektedir (Khaled vd., 1998). k parametresi aynı zamanda kaç tane küme merkezinin olacağı bilgisini taşıdığı için, K-means algoritması kullanılırken k parametresinin önceden girilmesi gerekmektedir (Bilgin ve Çamurcu, 2005). Kümeler belirlenirken; grup elemanların grup içi benzerliklerin en fazla ve gruplar arası benzerliklerin en az olması hedeflenmektedir (Demiralay ve Çamurcu 2005). Benzer olan kümeler oluşturulurken veriler arasındaki uzaklıklara bakılmaktadır. Bunun için eldeki veri kaç kümeye ayrılmak isteniyor ise o kadar küme merkezi belirlenmektedir (k parametresi) ve her elemanın bu belirlenen küme merkezlerine olan uzaklıkları tek tek hesaplanmaktadır. Hesaplamama sonucu her bir veri, kendine en yakın olan küme merkezine dâhil edilmektedir. Oluşan yeni kümelerin, küme merkezleri yeniden hesaplanmaktadır ve her bir elemanın yeni küme merkezlerine olan uzaklıkları da yeniden hesaplanmaktadır. Bu işlem hiçbir elemanın dâhil olduğu küme değişmeyinceye kadar devam etmektedir. En sonunda hiçbir elemanın dahil olduğu küme merkezi değişmiyor ise o zaman hesaplama durdurulmaktadır ve her eleman en son bulunduğu kümeye atanarak algoritma bitirilmektedir (Tekin vd., 2011).

Algoritmada uzaklık hesaplamalarından sonra birbirine en yakın olan elemanlar en çok benzerlik gösteren elemanlardır ve aralarında benzer oldukları tespit edilen elemanlar bir kümede toplanmaktadır (Çalışkan ve Soğukpınar, 2008).

K-means'ın en büyük eksiği verilerin kaç kümeye ayrılacağını, yani k parametresinin kaç olacağını tespit edememesidir. Bunun için, eğer verilerin kaç kümeden oluştuğu biliniyorsa k parametresi olarak bilinen küme sayısı girilebilmekte ya da farklı k parametreleri girilerek en uygun olanı tespit edilebilmektedir.

3.2.2. X-Means

K-means, kümeleme algoritmaları arasında oldukça popülerdir. K-means'ın çok fazla tercih ediliyor olmasının sebebi kolay anlaşılabilir bir yapıya sahip olmasının yanında model başarı oranının oldukça yüksek olmasıdır. Diğer bir açıdan popüler olmasının yanında bazı eksikleri mevcuttur (Akçapınar vd., 2016). Bu eksiklerden en belirgin olanı; küme sayısını ifade eden k parametresinin kullanıcı tarafından sabit bir değer olarak sağlanması gerektiğidir. Küme sayısının sabit bir k değeri ile sınırlandırılması küme sayısı için daha iyi olabilecek diğer alternatiflerin göz ardı edilmesi anlamına gelmektedir (Pelleg ve Moore, 2000). X-means'te veriler bir kd-tree'ye yerleştirilerek her aşamadaki istatistiksel veriler depolanmaktadır. Depolanan istatistiksel veriler içinde, belirli bir bölge için göz önünde bulundurulması gereken merkezlerin listesi de tutulmaktadır (Pelleg ve Moore, 2000). Böylece tüm alternatifler değerlendirilerek en iyi olanın seçilmesi imkânı sağlanmış olmaktadır (Muhr ve Granitzer, 2009). X-means algoritması, K-means algoritmasının geliştirilmiş hali olarak (Tsai ve Chiu, 2008) 2000 yılında Pelleg ve Moore tarafından geliştirilmiş bir kümeleme (bölütleme) algoritmasıdır (Pelleg ve Moore, 2000). Temelinde K-means algoritmasının çalışma şeklini kullanmasının yanı sıra eksik görülen yönünün tamamlanması amacıyla oluşturulmuştur (Kalogeratos ve Likas, 2012). X-means ile K-means'ın eksikliği olarak görülen küme sayısının algoritma tarafından belirlenememesi sorununa çözüm sunulmaktadır (Akçapınar vd., 2016). X-means'te küme sayısı için sabit bir değer belirtmek yerine makul olabilecek bir aralık belirtilmektedir. Böylece X-means belirtilen bu aralıktaki değerlerden (Steinley, 2006) en doğru oluşuna karar verdiği küme sayısını kendisi belirleyebilmektedir (Bholowalia ve Kumar, 2014). X-means yapısı içerisinde aşamalı olarak K-means algoritması çalıştırılmaktadır (Feng ve Hamerly, 2007). Alt kümelerin daha doğru oluşturulması için K-means her çalıştıktan sonra, hangi merkezlerin kendilerini bölmeleri gerektiği konusunda yerel kararlar oluşturularak ilerlenmektedir. Bu bölünme kararları bayesin bilgi kriteri (BIC) hesaplanarak verilmektedir (Jain, 2010; Hamerly ve Elkan, 2004). Her aşamada mevcut merkezlerin (ebeveyn) ve yeni oluşturulan yavru (çocuk) merkezlerin en iyi sonuçları blaclisting'e kaydedilmektedir. Kayıtlar tutulurken küme merkezlerinin yanında her k parametresi için model seçim kriterini puanlamak için kullanılan değerlerde tutulmaktadır (Pelleg ve Moore, 2000). Böylece

merkez konumlarının dikkatlice analiz edilmesine imkan sağlanmaktadır (Tucker vd., 2010). Temelde, algoritma verilen aralığın alt sınırına eşit k ile başlamakta ve üst sınıra ulaşılan kadar aşamalı olarak yeni merkezler eklemeye devam etmektedir. Bu işlemler sırasında, en iyi puanı elde eden centroid seti kaydedilmektedir. Bir sonraki işlemde elde olan puan, sistemde kayıtlı olan puandan daha yüksek ise yeni puan listeye eklenmektedir. Böylece liste sürekli olarak güncellenmektedir. Güncelleme prosedürü özyinelemelidir ve bir bölge içindeki noktalar için olası centroidlerin listesini tutmaktadır. Görevi, söz konusu bölgelerin merkez noktalarını uygun değerler ile güncellemektir. İlk olarak, girilen k parametre listesinin en küçük değerindeki sayı kadar rasgele oluşturulan merkez noktalarının yazılması ile başlamaktadır (Pelleg ve Moore, 2000). Yeni değerler hesaplandıktan sonra daha iyiler bulunduğça güncellenmektedir. En sonunda en yüksek değere sahip merkezler çıktı olarak kullanılmaktadır (Bholowalia ve Kumar, 2014).

3.3. Ön İşleme Teknikleri

Veri madenciliği yöntemleri ile analiz edilmek için toplanan veriler eksik olabilmektedir. Bu gibi durumlarda analizleri eksik veri ile gerçekleştirmek mümkündür. Aynı zaman da veri setinin kalitesini artırmak için eksik verilerin düzeltilmesi de tercih edilmektedir. Analizlerin eksiksiz veri üzerinde gerçekleştirilmesi kullanılan yöntemin başarı oranını arttırmada da oldukça önemli bir paya sahiptir. Eksik verinin düzeltilmesi veri ön işleme aşamasında yer almaktadır. Eksik veriyi düzeltmek için eksik verinin olduğu kayıtların çıkarılması (Sezgin ve Çelik, 2013), veya çeşitli yöntemlerle tamamlanması mümkündür (Aydın, 2007). Eksik olan verilerin tamamlanması için literatürde kabul görmüş birçok tamamlama yöntemi mevcuttur. Eksik değerler tamamlanırken dışardan genel bir sabit değer kullanılabilen, diğer verilerin özellikleri dikkate alınarak bir değer atanabilen (ortalama, mod ya da medyan gibi) (Aydın, 2007) ya da tahminlemeler sonucu bulunan değerler kullanılabilen (Regresyon analizi, Hot Deck, Naive Bayes İle Değer Atama, Son Gözlemi İleri Taşıma, Karar Ağaçlar, Beklenti Maksimizasyonu, Çoklu atama gibi) (Sezgin ve Çelik, 2013). Bu çalışmada eksik verilerin tamamlanması için diğer verilerin özellikleri dikkate alınarak bir değer ataması yapılmıştır. Veri seti içerisindeki her bir özneliğin frekansına bakılmış ve eksik olan veriler tamamlanırken frekansı en yüksek olan değer kullanılmıştır.

Günümüzde, veriler içinden anlam çıkarmak için veri madenciliği yöntemleri oldukça fazla kullanılmaktadır. Veri madenciliği ile bilginin keşfi süreci; veri seçimi ve toplanması, veri ön işleme, veri dönüştürme, veri madenciliği yöntemlerini uygulama ve en sonunda sonuçları değerlendirme şeklinde gerçekleşmektedir (Sebik ve Bülbül, 2018). Elde edilen verilerin kalitesi kullanılan yöntemin başarı oranını arttırmada çok büyük bir paya sahiptir (Oğuzlar, 2003). Verinin kalitesinin artırılması için kullanılan birçok veri ön işleme tekniği vardır (Mackinnon ve Glick, 1999). Veri setindeki eksik verilerin tamamlanması (Sezgin ve Çelik, 2013), gürültülü verilerin düzeltilmesi veya çıkarılması (Akgöbek ve Çakır, 2009), özneliklerin önem derecesinin belirlenmesi (Çetişli, 2006), bazı öznelikler üzerinde normalizasyon işleminin yapılması (Jayalakshmi ve Santhakumaran, 2011) kullanılan ön işleme tekniklerinden bazılarıdır. Bu çalışmada ön işleme tekniklerinden eksik verilerin tamamlanması ve normalizasyon kullanılmıştır. Literatürde kabul görmüş birçok normalizasyon çeşidi mevcuttur. Z-Score Normalizasyonu, Min-Max Normalizasyonu, Medyan Normalizasyonu, Sigmoid Normalizasyonu kabul gören normalizasyon yöntemlerinden bazılarıdır (Yavuz ve Deveci, 2012). Normalizasyon yöntemlerinin biri ya da birkaçı aynı anda kullanılabilir. Bu çalışmada Min-Max normalizasyon yöntemi kullanılmıştır.

4. UYGULAMA

4.1. Otizm Spektrum Bozukluğu Veri Seti

Bu çalışmada güncel ve gerçek OSB verilerinin çocuklar için olan alt kümesi kullanılmıştır. Verilerin ana kaynağı <http://fadifayez.com/autism-datasets/> internet sitesidir. Aynı zamanda Fadi Thabtah'ın 2017 yılında yaptığı bilimsel yayınlı birlikte UCI machine learning repository sitesinde paylaşılmış ve kullanmak isteyenler için erişime açılmıştır. Kullanılan veri setinin adı Autistic Spectrum Disorder Screening Data for Children Data Set'tir. Veri seti, OSB'nin teşhisi için literatürde kabul gören son parametrelere göre hazırlanan soruların cevapları ile oluşturulmuştur. Cevaplar ASDTests uygulaması ile toplanmıştır. Uygulamanın içinde yaş gruplarına göre ayrı kategoriler mevcuttur. Her kategoride on soru bulunmaktadır ve her biri kullanıcıların uygun cevabı dikkatlice seçmelerini sağlamak için bir resimle ilişkilendirilmiştir. Katılımcılara, bilgilerinin isimsiz tutulacağı ve sadece araştırma amacıyla paylaşılacağı bildirilmiştir. Değerlendirme tamamlamadan önce, katılımcılardan

araştırmanın amacını, gizlilik politikasını ve verilerin kullanımını açıklayan bir bilgilendirme yapılmıştır. Kullanılan veri setinde girdi olarak kullanılan öznitelikler arasında otizm spektrum bozukluğunu ölçümleyen on sorunun yanında demografik özellikleri içeren bilgilerde yer almaktadır. Toplamda 1100 kişiye uygulanan bu test, üç alt grup verilerini içermektedir. Gruplar çocuk, genç ve yetişkinler şeklindedir. Bu grupların oluşturulmasının sebebi ise OSB'nin teşhisi için sorulan bazı soruların gruplar arasında farklılık göstermesidir. Oluşturulan veri setinin çocuklar için olan alt kümesinde toplamda 292 örnek vardır ve bu çalışmada yalnızca çocuklar için olan küme kullanılmıştır. Veri seti sayısal ve kategorik veri türlerini içermektedir. Veri seti içerisindeki bazı örneklerde eksik veriler bulunmaktadır. Toplamda 21 öznitelikten oluşan veri setinde; 20 öznitelik bireylerin genel ve sağlık durumu ile ilgili bilgiler içerdiği için girdi olarak kullanılmış, 1 öznitelik ise bireyin otizmlili olma durumunu içerdiği için çıktı olarak kullanılmıştır. Girdi olarak kullanılan öznitelik parametrelerinden biri olan 'tarama yönteminin türü' bireyin yaş aralığının hangi grupta olduğunu ifade etmektedir. Yani otizm tarama testini gerçekleştiren bireyin çocuk, genç ya da yetişkin gruplarının hangisine dâhil olduğu bilgisini göstermektedir. Bu çalışmada sadece çocuklar için olan alt küme kullanılmıştır. Tüm veri seti çocuklara ait olduğu için grup bilgisine ihtiyaç duyulmamıştır. Bu sebepten dolayı veri seti içerisinde bu öznitelik çıkarılmıştır. Böylece girdi parametresinin sayısı 19' a düşürülmüştür. Özniteliklerin türü ve soruların detayları Tablo 1'de detaylı olarak verilmiştir.

Tablo 1

Veri setindeki özniteliklerin detayları

Öznitelik numarası	Öznitelik	Veri türü	Açıklama
1	Soru 1'in cevabı	Binary (0, 1)	Diğerleri duymadığında genellikle küçük sesler fark eder.
2	Soru 2'in cevabı	Binary (0, 1)	Genelde küçük detaylardan ziyade resmin tamamına odaklanır.
3	Soru 3'in cevabı	Binary (0, 1)	Bir sosyal grupta, birkaç farklı insanın konuşmasını kolayca takip edebilir.
4	Soru 4'in cevabı	Binary (0, 1)	Farklı etkinlikler arasında gidip gelmeyi kolay buluyor.
5	Soru 5'in cevabı	Binary (0, 1)	Adaylarla sohbete nasıl devam edeceğini bilmiyor.
6	Soru 6'in cevabı	Binary (0, 1)	Sosyal sohbetinde iyidir.
7	Soru 7'in cevabı	Binary (0, 1)	Bir hikaye okuduğunda, karakterin niyetlerini veya duygularını çözmekte zorlanır.
8	Soru 8'in cevabı	Binary (0, 1)	Okul öncesi eğitimindeyken, diğer çocuklarla oyun oynamayı sevmekten zevk alır.
9	Soru 9'in cevabı	Binary (0, 1)	Birinin ne düşündüğünü veya hissettiğini, sadece yüzlerine bakarak kolayca anlayabilir.
10	Soru 10'in cevabı	Binary (0, 1)	Yeni arkadaşlar edinmekte zorlanıyor.
11	Yaş	Number	Bireyin yıl olarak kaç yaşında olduğu bilgisi.
12	Cinsiyet	String	Bireyin erkek ya da kadın olduğu bilgisi.
13	Etnik köken	String	Bireyin hangi etnik kökene sahip olduğu bilgisi.
14	Sarılık ile doğmuş olmak	Boolean (evet-hayır)	Bireyin sarılık ile doğup doğmama bilgisi.
15	PDD'li aile üyesi	Boolean (evet-hayır)	Bireyin herhangi bir aile üyesinin PDD'sinin olup olmadığı bilgisi.
16	İkamet edilen ülke	String	Bireyin ikamet ettiği ülkenin bilgisi.
17	Daha önce tarama uygulamasının kullanılması	Boolean (evet-hayır)	Kullanıcının daha önce bir tarama uygulaması kullanıp kullanmadığı bilgisi.
18	Puanlama sonucu	Integer	Nihai puan, kullanılan tarama yönteminin puanlama algoritmasına dayanarak elde edilmiştir.
19	Teşti kim tamamlıyor	String	Bireyin testini kimin gerçekleştirdiği (Veli, kendisi, bakıcı, sağlık personeli, klinisyen, vb.) bilgisi.
20	Otizmlili olma durumu (Çıktı)	Boolean (evet-hayır)	Bireyin otizmlili olma durumunun bilgisi

(Thabtah vd., 2018)

Veri setinde bulunduğu halde veri setinin hangi yaş kategorisine uygulandığı bilgisini taşıyan ve tüm bireylerin çocuk olduğunu ifade etmek için 1 ile doldurulmuş olan "tarama yöntemi türü" özniteliği veri setinden çıkarılmıştır. Boolean türünde olan bireyin otizmlili olup olmama durumu ise sınıf etiketi olarak kullanılmıştır.

Çalışmada öncelikle öncelikle tüm öznitelikler sayısal değerlere çevrilmiştir. Bir sonraki aşamada veri seti üzerinde daha etkin analiz yapabilmek için eksik verilerin tamamlanması, verilerin 0 ile 1 arasında normalize edilmesi gibi veri ön işleme aşamaları gerçekleştirilmiştir. Daha sonra veriler yapay sinir ağları ve DKSBS ile sınıflandırılmış, k-means ve x-means ile kümelendirilmiştir. Tasarlanan ileri beslemeli YSA için kullanılan nöron sayılarının başarı oranına etkisi incelenmiş ve farklı yapıda düğümleri olan modeller test edilmiştir. Sınıflandırma yöntemlerinin model başarı oranını test etmek için veriler

%70 eğitim, %30 test olacak şekilde ayrılmıştır. Sonuçlara bakıldığı zaman veriler hem yapay sinir ağları hem de DKSBS ile sınıflandırıldığında başarı oranını hem test hem de eğitim setinde %100 olduğu görülmüştür. Kümeleme yöntemleri ön öğreticisiz yöntemlerdir. Dolayısı ile kümelenen veriler eğitim ve test olarak ikiye ayrılmamıştır ve verilerin tamamı eğitim için kullanılmıştır. Kümelenen verilerin başarı oranı k-means de % 89,73, x-means de %88,0 şeklindedir

4.2. Bulgular ve Değerlendirme

Literatür incelendiğinde otizm konusunda veri madenciliği yöntemleri kullanılarak gerçekleştirilen çok fazla çalışma olduğu görülmektedir. Thabtah (2017), çalışmasında bu konuya dikkat çekmiş, gerçekleştirilen çalışmalarda oldukça güzel sonuçların elde edilmiş olmasıyla birlikte kullanılan verilerin güncelliğini yitirdiğinden bahsetmiştir. Bu sebepten dolayı da güncel verilerle çalışılması gerektiğini vurgulamıştır. Devamında ise Thabtah (2018), yeni bir çalışma ile güncel ölçeğe uygun şekilde toplanan verileri paylaşmıştır. Bu çalışmada da Thabtah (2018), tarafından güncel ölçeğe uygun şekilde toplanan ve yapılan bilimsel yayın ile paylaşılan verilerin çocuklar için olan alt kümesi kullanılmıştır. Kullanılan veri seti 292 örnekten oluşmaktadır ve veri setinde toplamda 21 öznitelik mevcuttur. Veriler çok eski olmadığından bu veriler ile yapılan çalışmalarda sınırlıdır ve bu çalışma ile elde edilen bulgular diğer çalışmalarla karşılaştırılmıştır.

Veri setinin orijinal hali sayısal ve string ifade türünde değişkenler içermektedir. Bu çalışma kapsamında kullanılan normalizasyon işlemi ve sınıflandırma işlemi için kullanılan yöntemler ise string ifadelerle çalıştırılmamaktadır. Bu sebepten dolayı öncelikle tüm veri seti sayısal değerlere dönüştürülmüştür. Bu dönüşüm yapılırken string ifadelerle etiketlenmiş olan kategorik veriler, sayısal ifadeler ile etiketlenecek şekilde değiştirilmiştir. Örneğin bayan ya da erkek olarak etiketlenen kategorik veriler 0 ve 1 ile değiştirilmiştir. Etnik köken, ikamet edilen ülke gibi değerler de aynı şekilde 1'den başlanarak sıralı olarak etiketlenmiştir.

Tüm veriler sayısal hale getirildikten sonra her bir özneliğin kendi arasında frekansı çıkarılmıştır. Veri seti içerisindeki boş değerler frekansı en yüksek olan değer ile doldurulmuştur. Devamında ise birçok normalizasyon yöntemi denemiş en yüksek başarı oranına sahip literatürde de oldukça fazla kullanılan Min-Max normalizasyon yöntemi ile elde edilen değerler oluşturulmuştur. Yöntemlerin başarılarının karşılaştırılması için veri setinin bir kısmı eğitim bir kısmı test verileri olarak ayrılmıştır. Bu oranın yüzde kaç olarak ayrılması gerektiği konusunda literatür incelendiğinde çok farklı oranların kullanılmış olduğu görülmekle birlikte, en yaygın oran %70 eğitim verisi, %30 test verisi olarak ayrıldığı görülmektedir. Genel yaklaşıma uyması açısından bu çalışmada da %70'e karşılık gelen ve rastgele olarak belirlenen 205 örnek eğitim, %30'a karşılık gelen ve rastgele olarak belirlenen 87 örnek ise test için kullanılmıştır. Veriler YSA ve DKSBS ile sınıflandırılmış, K-Means ve X-Means ile kümelenebilir.

4.2.1. Yapay Sinir Ağı ile Elde Edilen Sonuçlar

Tasarlanan ileri beslemeli YSA ile kullanılan nöron sayılarının başarı oranına etkisi incelenmiş ve farklı yapıda düğümleri olan modeller test edilmiştir. YSA'nın performansı Tablo 2 de verilmiştir

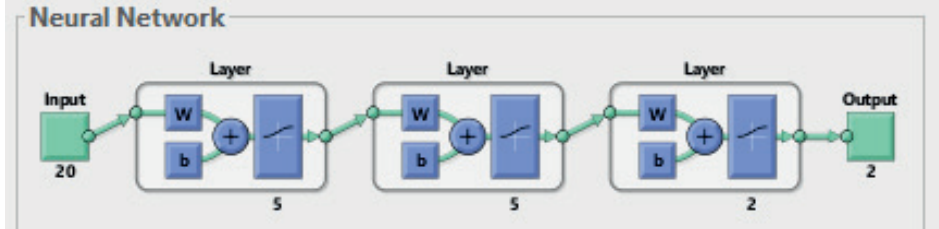
Tablo 2

Yapay Sinir Ağı'nın performansı

Giriş katmanındaki nöron sayısı	1.Gizli katmandaki nöron sayısı	2.Gizli katmandaki nöron sayısı	Çıkış katmanındaki nöron sayısı	Eğitim verisindeki doğruluk oranı	Test verisindeki doğruluk oranı
20	1	1	2	%91,21	%90,80
20	10	10	2	%100	%97,70
20	5	5	2	%100	%98,85
20	30	30	2	%100	%96,56
20	15	15	2	%100	%96,56
20	50	50	2	%100	%97,70
20	20	20	2	%100	%96,56

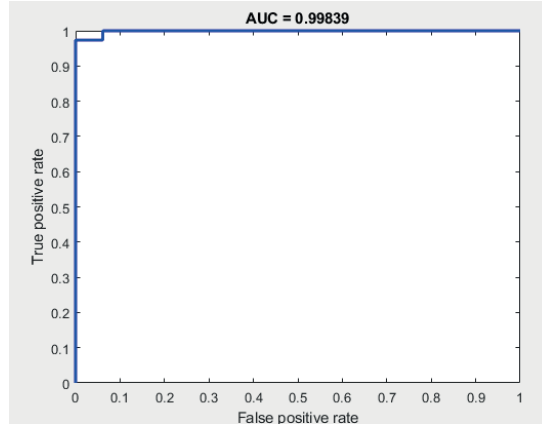
Tasarlanan modellerin başarı oranına bakıldığı zaman oldukça yüksek oranda başarıya ulaşıldığı görülmüştür. Farklı nöron sayıları ile tasarlanan modeller içerisinde en yüksek performans giriş katmanının da 20, iki gizli katmanda 5 ve çıkış katmanında 2 nöron kullanılarak tasarlanan modelde görülmüştür. Eğitim setinde %100 test setinde ise %98.85 oranında doğrulukla

başarı elde edilmiştir. Bu durum eğitim setindeki tüm örneklerin doğru sınıflandırıldığı, test veri setinde ise yalnızca bir örneğin yanlış sınıflandırıldığı anlamına gelmektedir. Tablo 4’de görüldüğü gibi gizli katmandaki nöron sayısını arttırmak başarı oranının artması ile doğru orantılı değildir. Nöron sayısını arttırmak bazı veri setleri için modelin başarı oranını arttırmakla beraber, bazı veri setleri için modelin başarı oranını düşürmektedir. Dolayısı ile her zaman nöron sayısının çok fazla ya da çok az kullanılması ile alakalı sabit bir yaklaşımı benimsemek yerine birkaç alternatif belirlenerek en başarılı olan modelin tercih edilmesi başarı oranını arttırabilmektedir. Başarı oranının en yüksek olarak görüldüğü YSA modeli Şekil 1de gösterilmiştir.



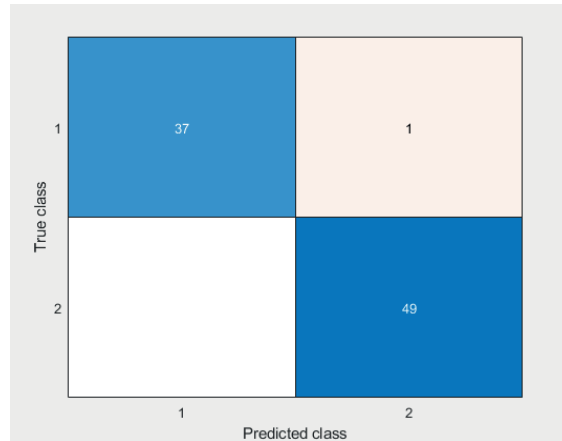
Şekil 1. Başarı oranının en yüksek olarak görüldüğü YSA modeli

Eğitilen ağın test verisi için çizdirilen ROC eğrisi (Receiver Operating Characteristic) Şekil 2 de verilmiştir.



Şekil 2. Eğitilen ağın test verisi için çizdirilen ROC eğrisi

Şekil 2 de görüldüğü üzere ROC eğrisi altında kalan alan (AUC) değeri 1’e çok yakın çıkmıştır. Bu durum, test verisinde yapılan sınıflandırmada oldukça yüksek bir başarı sağlandığını göstermektedir. Eğitilen ağın test verisi için çizdirilen hata matrisi Şekil 3 de verilmiştir.

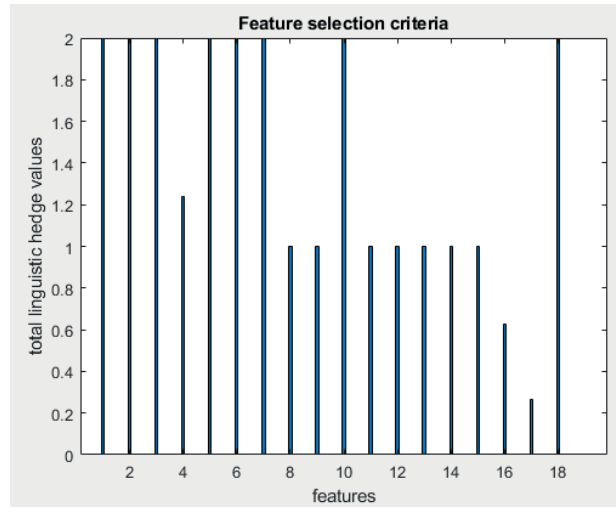


Şekil 3. Eğitilen ağın test verisi için çizdirilen hata matrisi

Şekil 3 de görüldüğü gibi eğitilen ağın test veri seti için gerçekleştirdiği sınıflandırmanın çok yüksek bir kısmı doğru tahminlenmiş, yalnızca bir örnek yanlış sınıflandırılmıştır. Bu durum test verisi için oldukça yüksek bir başarı anlamı taşımaktadır. Ağın test verisi için ürettiği hata değerlerine detaylı olarak bakıldığında MSE değerinin $4,822e-04$; RMSE değerinin ise $0,022$ olduğu görülmektedir. Eğitilen ağın eğitim verileri için MSE ve RMSE değerlerinden de anlaşıldığı gibi ağın hata değerleri oldukça düşüktür. Eğitilen ağın, test verileri için gerçekleştirdiği Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri sırası ile $0,989$ $0,974$ 1 $0,987$ şeklindedir. Bu durum eğitilen ağın test verileri içinde oldukça başarılı sonuçlar verdiğini göstermektedir.

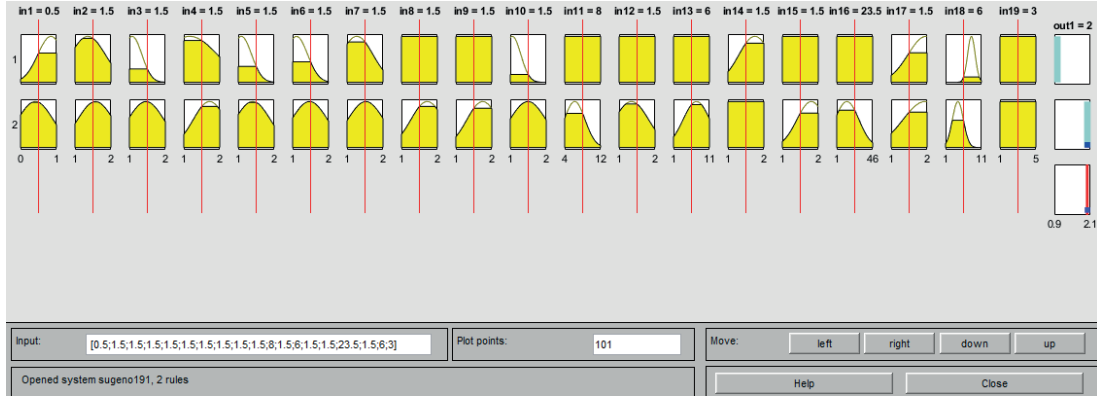
4.2.2. Dilsel Kuvvetli Sinir-Bulanık Sınıflayıcı ile Elde Edilen Sonuçlar

DKSBS ile öncelikle öznelik seçimi gerçekleştirilmiş devamında ise sınıflandırma işlemi yapılmıştır. Yapılan sınıflandırmada veriler %70 eğitim, %30 test verisi şeklinde ayrılarak gerçekleştirilmiştir. Sınıflandırma sonuçlarına bakıldığı zaman eğitim ve test verilerinde %100 oranda başarı elde edildiği görülmüştür. Yöntemin eğitim verileri için hesaplanan hata değerlerine bakıldığı zaman MSE değerinin $3,099e-32$, RMSE değerinin ise $1,760e-16$ olduğu görülmektedir. Gerçekleştirilen öznelik seçimi, özneliklerin önem düzeyine bakılarak belirlenmiştir. Özneliklerin DKSBS tarafından belirlenen önem düzeyleri Şekil 4 de verilmiştir.



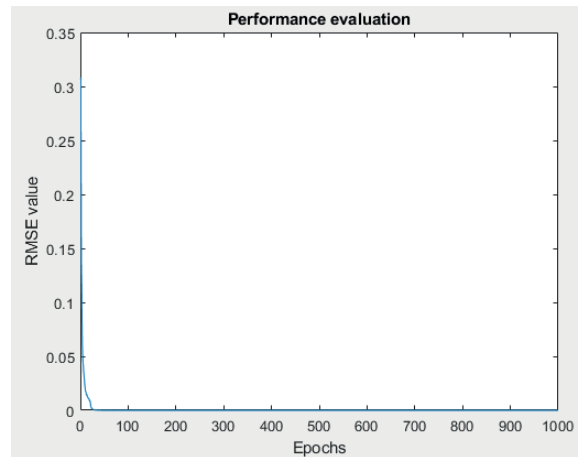
Şekil 4. Özneliklerin DKSBS tarafından belirlenen önem düzeyleri

Şekil 4 de numaralandırılmış özneliklerin sırası Tablo 1 deki sıra ile aynı olacak şekilde 1'den 19'a kadar sıralı olarak ilerlemiştir. Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın sınıflandırma için öznelikler üzerinde belirlemiş olduğu önem derecelerine bakıldığı zaman beş farklı önem düzeyine sahip grupların oluştuğu görülmektedir. Bu grupları en büyük öneme sahip onları 5 olacak şekilde, 5 den 1'e kadar derecelendirilecek olursa; 5 önem düzeyine sahip öznelikler 1,2,3,5,6,7,10,18 numaralı olanlar, 4 önem düzeyine sahip öznelik 4 numaralı olan, 3 önem düzeyine sahip öznelikler 8,9,11,12,13,14,15 numaralı olanlar, 2 önem düzeyine sahip öznelik 16 numaralı olan ve 1 önem düzeyine sahip olan 17 numaralı olan özneliktir. 19 numaralı özneliğin ise önem derecesi 0 olarak belirlenmiştir. Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın ifade ettiği özneliklerin katkı dereceleri Şekil 5 de görülmektedir.



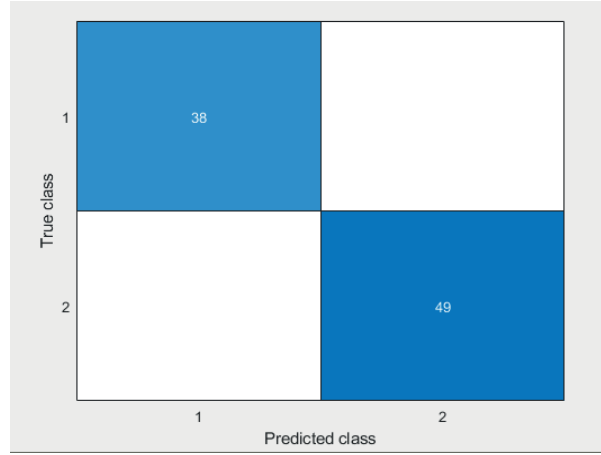
Şekil 5. Dilsel kuvvetli sinir-bulanık sınıflayıcı'ya ait bulanık kuralların modeli

Tablo 1 de yer alan 1,2,3,5,6,7,10 ve 18 numaralı öznelikler olan çocuğun; diğerleri duymadığında genellikle küçük sesleri fark edip etmediği, genelde küçük detaylardan ziyade resmin tamamına odaklanıp odaklanmadığı, bir sosyal grupta, birkaç farklı insanın konuşmasını kolayca takip edebilip edemediği, adaylarla sohbet nasıl devam edeceğini bilip bilmediği, sosyal sohbetinde iyi olup olmadığı, bir hikaye okuduğunda, karakterin niyetlerini veya duygularını çözmekte zorlanıp zorlanmadığı, yeni arkadaşlar edinmeyi zor bulup bulmadığı ve kullanılan tarama yönteminin puanlama algoritmasına dayanarak elde edilmiş nihai puanın yer aldığı öznelikler sınıflandırma üzerinde en çok etkiye sahip olan öznelikler olarak belirlenmiştir. İkinci en etkili öznelik olarak 4 numara ile ifade edilen; farklı etkinlikler arasında gidip gelmeyi kolay bulup bulmadığı durumu belirlenmiştir. Üçüncü en etkili öznelikler 8,9,11,12,13,14,15 numaralarıyla ifade edilen; okul öncesi eğitimindeyken, diğer çocuklarla oyun oynamayı sevip sevmediği, birinin ne düşündüğünü veya hissettiğini, sadece yüzlerine bakarak kolayca anlayabilip anlayamadığı, bireyin yıl olarak kaç yaşında olduğu bilgisi, bireyin erkek ya da kadın olduğu bilgisi, bireyin hangi etnik kökene sahip olduğu bilgisi, bireyin sarılık ile doğup doğmama bilgisi ve bireyin herhangi bir aile üyesinin PDD'sinin olup olmadığı bilgisi'dir. Dördüncü en etkili öznelik 16 numara ile ifade edilen; bireyin ikâmet ettiği ülkenin bilgisidir. Beşinci en etkili öznelik 17 numara ile ifade edilen; kullanıcının daha önce bir tarama uygulaması kullanıp kullanmadığı bilgisidir. Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın eğitim performansı Şekil 6 da gösterilmiştir.



Şekil 6. Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın eğitim performansı

Şeki 6 da görüldüğü gibi dilsel kuvvetli sinir-bulanık sınıflayıcı'nın hata değerleri oldukça düşüktür. Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın eğitildikten sonra test verisi için çizdirilen hata matrisi Şekil 7 de verilmiştir.



Şekil 7. Test verisi için çizdirilen hata matrisi

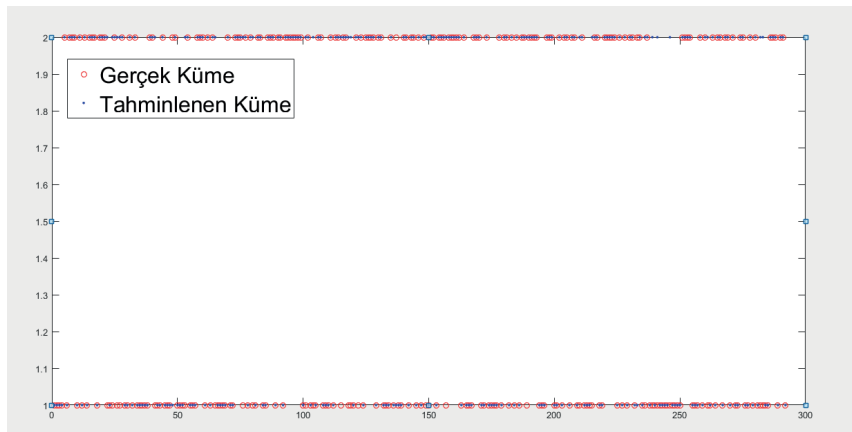
Şekil 7 de görüldüğü gibi eğitilen dilsel kuvvetli sinir-bulanık sınıflayıcı'nın test veri seti için gerçekleştirdiği sınıflandırmanın tamamı doğrudur.

Eğitilen dilsel kuvvetli sinir-bulanık sınıflayıcı'nın, test verileri için gerçekleştirdiği Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri ise sırası ile 1,1,1,1 şeklindedir. Bu durum eğitim verilerinin tamamının doğru sınıflandırıldığını göstermektedir.

4.2.3. K-Means ile Elde Edilen Sonuçlar

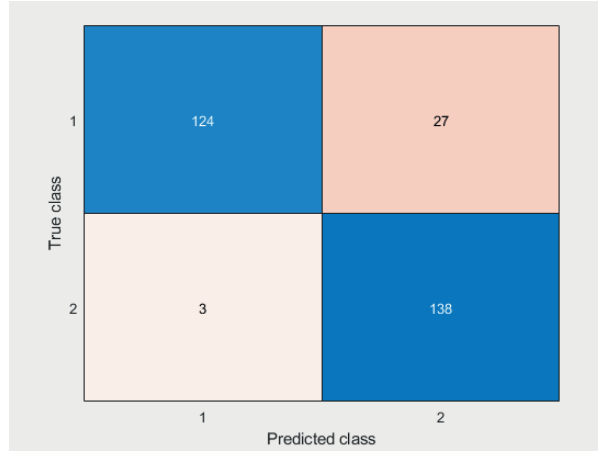
Kümeleme algoritmaları ön öğreticisiz yöntemlerdir ve verilerin sınıfları önceden belli değildir. Bu sebepten dolayı veriler eğitim ve test verisi olarak ayrılmamış, tamamı eğitim için kullanılmıştır.

K-means algoritmasında k parametresi ile ifade edilen küme sayısı önceden belirlenmeli ve çalıştırılmadan önce algoritmaya verilmelidir. Bu çalışmadaki verilerin OSB olup olmadığını gösterecek şekilde iki sınıfı vardır. Dolayısı ile kümeler oluşturulurken iki küme olması istendiğinden k parametresi iki olarak girilmiş ve veriler ikiye kümelendi. K-means algoritması ile kümelenen verilerin sınıfları, gerçek küme sınıfları ile karşılaştırıldığında toplamda 292 veriden 262'si doğru sınıflandırılmış ve sınıflandırmanın doğruluk oranı % 89,73 olmuştur. Gerçek sınıflar ve K-means ile elde edilen sınıflar Şekil 8 de verilmiştir.



Şekil 8. Gerçek sınıflar ve K-means ile elde edilen sınıflar

K-means algoritması ile elde edilen sonuçlar için çizdirilen hata matrisi Şekil 9'da verilmiştir.

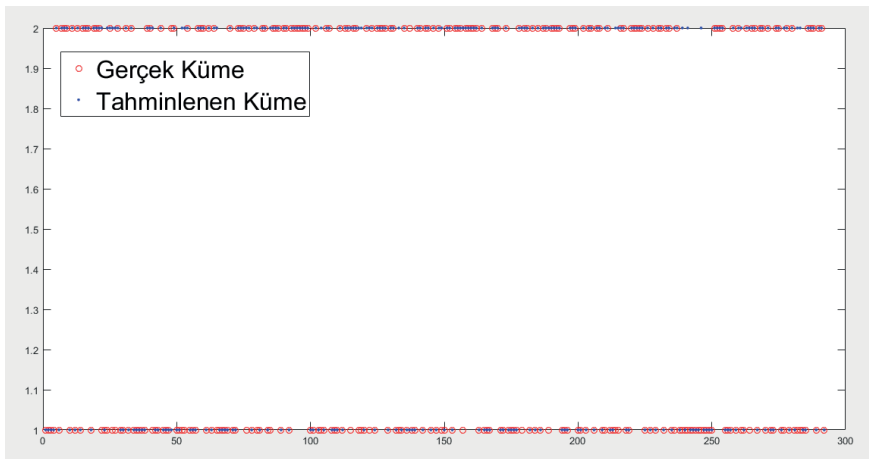


Şekil 9. K-means algoritması ile elde edilen sonuçlar için çizdirilen hata matrisi

Şekil 9’da görüldüğü gibi K-means algoritması ile kümelenen verilerin sınıfları çoğunlukla doğru tahminlenmiş olup, toplamda yanlış tahminlenen örnek sayısı 30 olmuştur. K-means algoritması ile kümelenen verilerin Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri sırası ile 0,897 0,821 0,979 0,892 şeklindedir.

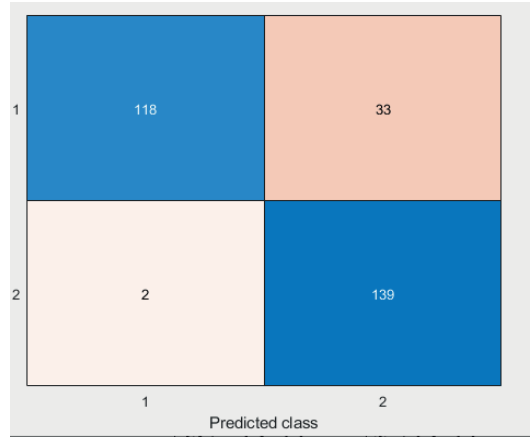
4.2.4. X-Means ile Elde Edilen Sonuçlar

X-means algoritması K-means algoritmasının geliştirilmesiyle oluşturulmuş daha güncel bir kümeleme yöntemidir. X-means, K-means’in çalışma yapısını kullanmasının yanında K-means’in geliştirilmiş halidir. Küme sayısının sabit bir değer olarak belirlenmesinin yerine bir aralık belirlenmesi ile çalışmaktadır. Böylece veri seti için en uygun olan küme sayısı algoritma tarafından belirlenebilmektedir. Küme sayısı belirlenirken, en iyi küme sayısını belirlemek için Bayesin bilgi kriteri kullanılmaktadır. Bu çalışmada, sınıf aralığı 2 ile 4 arasında girilerek çalıştırılan X-means algoritmasında küme sayısı algoritma tarafından 2 olarak belirlenmiştir. X-means algoritması ile ikiye kümelenen verilerin sınıfları, gerçek küme sınıfları ile karşılaştırıldığında toplamda 292 veriden 257’si doğru sınıflandırılmış ve sınıflandırmanın doğruluk oranı %88,02 olmuştur. Gerçek sınıflar ve X-means ile elde edilen sınıflar Şekil 10’da verilmiştir.



Şekil 10. Gerçek sınıflar ve X-means ile elde edilen sınıflar

X-means algoritması ile elde edilen sonuçlar için çizdirilen hata matrisi Şekil 11'de verilmiştir.



Şekil 11. X-means algoritması ile elde edilen sonuçlar için çizdirilen hata matrisi

Şekil 11'de görüldüğü gibi X-means algoritması ile kümelenen verilerin sınıfları çoğunlukla doğru tahminlenmiş olup, toplamda yanlış tahminlenen örnek sayısı 35 olmuştur. X-means algoritması ile kümelenen verilerin Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri sırası ile 0,880 0,781 0,986 0,871 şeklindedir.

4.2.5. Farklı Yöntemler ile Analiz Edilen Çocuklar İçin Otizm Spektrum Bozukluğu Verilerinin Tahminleme Başarılarının Karşılaştırılması

Bu çalışmada OSB veri setinin çocuklar için olan alt kümesi YSA ve DKSBS ile sınıflandırılmış, K-means ve X-means yöntemleri ile kümelendirilmiştir. Kullanılan yöntemlerin performans değerleri Tablo 2'de verilmiştir.

Tablo 2

Kullanılan yöntemlerin performans değerleri

Yöntem	Eğitim setinde doğruluk oranı	Test setinde doğruluk oranı
Yapay sinir ağları	%100	%98,85
Dilsel kuvvetli sinir-bulanık sınıflayıcı	%100	%100
K-Means	%89,73	Tüm veriler eğitim için kullanılmıştır
X-Means	%88,02	Tüm veriler eğitim için kullanılmıştır

Tablo 2'de görüldüğü gibi YSA ile sınıflandırma başarı oranı oldukça yüksek olmasına rağmen en yüksek başarı oranı DKSBS ile elde edilmiştir. DKSBS ile yapılan sınıflandırma da eğitim ve test verisi için %100 oranında doğruluk gerçekleştirilmiştir. Bu durum dilsel kuvvetli sinir-bulanık sınıflayıcının veri setindeki tüm örnekleri doğru sınıflandırdığı anlamını taşımaktadır. Kümeleme yöntemlerinin başarı oranları incelendiği zaman genel anlamda sınıflandırma yöntemlerine göre daha az başarılı olduğu ve K-means yöntemi ile elde edilen sonuçların, X-means yöntemine göre daha başarılı olduğu görülmektedir. Kullanılan Yöntemlerin Eğitim veri setleri için Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri Tablo 3'de verilmiştir.

Tablo 3

Kullanılan yöntemlerin eğitim veri setleri için Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri

Yöntem	Doğruluk	Duyarlılık	Belirleyicilik	F-Ölçütü
Yapay sinir ağları	1	1	1	1
Dilsel kuvvetli sinir-bulanık sınıflayıcı	1	1	1	1
K-Means	0,897	0,821	0,979	0,892
X-Means	0,880	0,781	0,986	0,871

Tablo 3'de sınıflandırma ve kümeleme yöntemlerinin eğitim verisi için hesaplanan Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri görülmektedir. Görüldüğü üzere sınıflandırma yöntemlerinde hesaplanan tüm değerler 1'dir. Kümeleme yöntemlerinde ise ortalama olarak K-means için hesaplanan değerler X-means'ten daha yüksektir.

Kullanılan Yöntemlerin test veri setleri için Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri Tablo 4’de verilmiştir.

Tablo 4

Kullanılan yöntemlerin test veri setleri için Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri

Yöntem	Doğruluk	Duyarlılık	Belirleyicilik	F-Ölçütü
Yapay sinir ağları	0.989	0.974	1	0.987
Dilsel kuvvetli sinir-bulanık sınıflayıcı	1	1	1	1
K-Means	Tüm veriler eğitim için kullanılmıştır	Tüm veriler eğitim için kullanılmıştır	Tüm veriler eğitim için kullanılmıştır	Tüm veriler eğitim için kullanılmıştır
X-Means	Tüm veriler eğitim için kullanılmıştır	Tüm veriler eğitim için kullanılmıştır	Tüm veriler eğitim için kullanılmıştır	Tüm veriler eğitim için kullanılmıştır

Tablo 4’de görüldüğü gibi test verisi için DKSBS ile yapılan sınıflandırmanın Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri 1 olarak hesaplanmıştır. Bu durum dilsel kuvvetli sinir-bulanık sınıflayıcının test verisinin tamamını doğru sınıflandırdığı anlamı taşımaktadır. Test verisi için YSA ile yapılan sınıflandırmanın Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri de 1’e oldukça yakın değerler çıkmıştır. Bu durum test verisi için dilsel kuvvetli sinir-bulanık sınıflayıcıdan daha başarısız olmakla birlikte kullanılabilir bir yöntem olduğunu ortaya koymaktadır. K-Means ve X-Means ile kümelene verilerin tamamı eğitim için kullanılmış, dolayısı ile test verisi için hesaplanan Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri bulunmamaktadır. Çalışmanın performans değerleri ve aynı veri setini kullanan diğer çalışmalardaki performans değerleri Tablo 5’de verilmiştir.

Tablo 5

Çalışmanın performans değerleri ve aynı veri setini kullanan diğer çalışmalardaki performans değerleri

Çalışma	Yöntem	Performans
Thabtah, 2018	Logistic regression	%97,94
Thabtah, 2018	Naive bayes	%92,80
Akyol ve Karaci, 2018	Lojistik regresyon	%92,00
Akyol ve Karaci, 2018	Bulanık kuralı lojistik regresyon kombinasyonu	%97,33
De Campos Souza ve Guimaraes, 2018	Bulanık sinir ağı mimarisi	%100
Hutchinson vd., 2019	J48 decision tree	%92,80
Hutchinson vd., 2019	Naive bayes	%96,20
Bu çalışma	Yapay sinir ağları	%98,85
Bu çalışma	Dilsel kuvvetli sinir-bulanık sınıflayıcı	%100
Bu çalışma	K-Means	%89,73
Bu çalışma	X-Means	%88,02

Tablo 5’de görüldüğü gibi aynı veri seti farklı çalışmalarda farklı yöntemlerle sınıflandırılmıştır ama bu çalışma dışında bir kümeleme işlemine rastlanmamıştır. Yapılan çalışmalardaki sonuçlar incelendiği zaman genel olarak başarı oranının %90’ın üzerinde olduğu görülmekle birlikte %100 olan bir çalışma da bulunmaktadır. Bu çalışma ile elde edilen sonuçlar ile diğer çalışmalarda elde edilen sonuçlar karşılaştırıldığı zaman ise verilerin DKSBS ve YSA ile sınıflandırıldığında sınıflandırma başarısının logistic regression, naive bayes, bulanık kuralı lojistik regresyon kombinasyonu ve j48 decision tree ile sınıflandırılan çalışmalardaki sınıflandırma başarısından daha yüksek olduğu görülmektedir. Bulanık sinir ağı mimarisi ile gerçekleştirilen sınıflandırma ve bu çalışma kapsamının da DKSBS ile gerçekleştirilen sınıflandırmanın başarı oranları ise aynıdır. İki yöntem ile sınıflandırılan verilerdeki doğruluk aynı orandadır ve bu oran %100 şeklindedir. Bu durum iki yöntemde tüm örnekleri doğru sınıflandırdığı anlamına gelmektedir. Bu çalışma kapsamında gerçekleştirilen kümeleme işlemi sonuçları incelendiği zaman ise kümeleme yöntemlerinden K-means yönteminin X-means yöntemine göre daha başarılı olduğu ve iki kümeleme yönteminin de başarı oranının sınıflandırma yöntemlerine göre daha düşük olduğu sonucuna ulaşılmaktadır.

5. TARTIŞMA, SONUÇ ve ÖNERİLER

Bu çalışma ile OSB konusunda veri madenciliği, yapay zeka gibi teknikleri kullanarak gerçekleştirilen çalışmalar incelenmiştir. Bu konuda oldukça başarılı sonuçlara ulaşan birçok çalışmanın olduğu görülmektedir. Bunun yanı sıra bilimsel bir yayın ile son dönemde gerçekleştirilen çalışmalarda güncelliğini yitiren veri setlerinin kullanıldığı eleştirisi dikkat çekmektedir. Yayının detayları incelendiği zaman güncel ve başarılı yöntemlerin kullanılarak teknolojik gelişmelerden faydalanıldığı

fakat deneysel çalışmalarda kullanılan verilerin güncelliğini yitirdiği vurgusu yapıldığı görülmektedir. Çalışmanın devamında ise eksiklik olarak görülen güncel olmayan veriler, geliştirilen son ölçeğe dayalı olarak toplanmış ve bir takım işlemler gerçekleştirildikten sonra paylaşılmıştır. Veri seti çocuk, genç, yetişkin şeklinde gruplandırılmak üzere toplamda 1100 örnek içermektedir. Bu çalışmada ise aynı veri setinin 292 örnek içeren çocuklar için olan alt küme kullanılmıştır. Toplamda 20 girdiden oluşan fakat tüm örneklerde aynı olduğu için çıkarılan bir öznelikten sonra 19'a düşürülen girdiler sonucunda bireyin otizmlili olma durumunu ifade eden çıktının tahminlenmesi için veri setine veri madenciliği yöntemlerinden olan sınıflandırma ve kümeleme işlemleri uygulanmıştır. Sınıflandırma ve kümeleme işlemleri uygulanmadan önce öncelikle tamamen numeric değerlere dönüştürülen veri seti üzerinde özneliklerin frekansına bakılarak eksik veriler tamamlanmış ve normalizasyon işlemleri gerçekleştirilmiştir. Yöntemlerin çalıştırılması için hazır hale getirilen veri seti son olarak %70 eğitim, %30 test verisi olmak üzere ayrılmıştır. Sınıflandırma işlemi için veriler test ve eğitim verisi olacak şekilde ayrılarak sınıflandırılmış, elde edilen sonuçlar birçok parametre ile ifade edilmiştir. Sınıflandırma işlemi için kullanılan yöntemler YSA ve DKSBS olmuştur. Kümeleme yöntemleri önöğreticisiz yöntemler olduğundan dolayı veri seti eğitim ve test olarak ayrılmamış tamamı eğitim verisi olacak şekilde kümeleneştir. Kümeleme işlemi için K-means ve X-means yöntemleri kullanılmıştır. Elde edilen sonuçlar birçok parametre ile ifade edilmiş olmakla birlikte tüm yöntemlerde ortak paydada karşılaştırma yapılabilmesi için temel olarak Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü kullanılmıştır.

Sınıflandırma sonuçlarına bakıldığı zaman eğitim veri seti için YSA ve DKSBS olmak üzere iki yöntemde başarı oranının %100 olduğu görülmüştür. Dolayısı ile iki yöntemin eğitim veri seti için hesaplanan Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü değerleri 1 olarak hesaplanmıştır. Bu durum iki yöntemde eğitim verilerinin tamamını doğru sınıflandırdığı anlamına gelmektedir. Test seti için elde edilen sonuçlara bakıldığı zaman başarı oranlarının birbirinden farklı olduğu görülmektedir. Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın test verisi için gerçekleştirdiği başarı oranı eğitim verisindeki başarı oranı ile aynı olmuştur. Dolayısı ile test seti için hesaplanan Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü değerleri 1 olarak hesaplanmıştır. Bu durum test verilerinin DKSBS tarafından tamamının doğru sınıflandırdığı anlamına gelmektedir. YSA'nın test verisi için gerçekleştirdiği başarı oranı ise %98.85 olmuştur. Test seti için hesaplanan Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü değerleri sırası ile 0,989 0,974 1 0,987 olarak hesaplanmıştır. Bu sonuçlar YSA ile yapılan sınıflandırmanın test verisinde 87 örnekten yalnızca 1 tanesini yanlış sınıflandırdığı geriye kalan 86 örneği doğru sınıflandırdığını anlamını taşımaktadır. Sınıflandırma yöntemlerinin başarılarına bakıldığı zaman iki yöntemde oldukça başarılı olduğu ve veri setini sınıflandırmak için kullanılabilmesi görülmekle birlikte, tüm verileri doğru sınıflandırdığından dolayı dilsel kuvvetli sinir-bulanık sınıflayıcının daha başarılı olduğu söylenebilmektedir.

Kümeleme sonuçlarına bakıldığı zaman eğitim ve test verisi olarak ayrılmadığı tüm verilerin eğitim için kullanıldığı görülmektedir. Dolayısı ile K-means ve X-means yöntemi ile kümelenen verilerde yalnızca eğitim verisi için sonuçlar bulunmaktadır. K-means ile kümelenen verilerin başarı oranı %89,73 olmuştur ve verilerin hesaplanan Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü değerleri sırası ile 0,897 0,821 0,979 0,892 şeklindedir. Bu sonuçlar K-means ile kümelenen verilerin 262 tanesinin doğru, 30 tanesinin yanlış sınıflandırıldığı anlamına gelmektedir. X-means ile kümelenen verilerin başarı oranı ise %88,02 olmuştur ve verilerin hesaplanan Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü değerleri sırası ile 0,880 0,781 0,986 0,871 şeklindedir. Bu sonuçlar X-means ile kümelenen verilerin 257 tanesinin doğru, 35 tanesinin yanlış sınıflandırıldığı anlamına gelmektedir. Bu çalışma kapsamında elde edilen sonuçlara göre kümeleme yöntemlerinin başarılarına bakıldığı zaman sonuçların genel anlamda yüksek olduğu fakat sınıflandırma yöntemlerine göre daha düşük kaldığı görülmekle birlikte bu çalışmadaki veri seti için K-means yönteminin X-means yöntemine göre daha başarılı sonuçlar verdiği tespit edilmiştir.

Bu çalışmada elde edilen sonuçlar ile aynı veri seti ile yapılan diğer çalışmalardaki sonuçlar karşılaştırıldığı zaman Thabtah (2018), logistic regression ile yaptığı sınıflandırmanın başarı oranı %97,94 naive bayes ile yaptığı sınıflandırmanın başarı oranı %92,80 bulmuştur. Akyol ve Karaci (2018), lojistik regresyon ile yaptıkları sınıflandırmanın başarı oranı %92,00 bulanık kurallı lojistik regresyon kombinasyonu ile yaptıkları sınıflandırmanın başarı oranı %97,33 olarak elde etmişlerdir. De Campos Souza ve Guimaraes (2018), bulanık sinir ağı mimarisi ile yaptıkları sınıflandırmanın başarı oranı %100 olarak tespit etmişlerdir. Hutchinson ve arkadaşları (2019), j48 decision tree ile yaptıkları sınıflandırmanın başarı oranını %92,80

naive bayes ile yaptıkları sınıflandırmanın başarı oranını %96,20 olarak bulmuşlardır. Bu durum genel olarak veri setinin sınıflandırmaya uygun bir veri seti olduğunu göstermektedir. Aynı zamanda bu çalışma ile %100 oranında başarıya ulaşan DKSBS yönteminin kullanılan veri seti için literatürdeki birçok yöntemden daha başarılı sonuçlar elde ettiğini de göstermektedir. Elde edilen başarı oranlarının yüksek olması önemli olmakla birlikte gerçek hayat problemlerinde %100 oranında başarının var olması gerçekleşmesi zor bir ihtimaldir ve hata payının her zaman olabileceği ihtimalinin göz önünde bulundurulması gerekmektedir.

Bu çalışmanın literatüre olan katkıları ise şöyledir; öncelikle aynı veri seti ile yapılan çalışmalar incelendiği zaman logistic regression, naive bayes, bulanık kurallı lojistik regresyon kombinasyonu, bulanık sinir ağı mimarisi, j48 decision tree gibi yöntemlerin kullanılması ile yapılan sınıflandırma işlemleri olmakla birlikte çocuklar için olan alt kümede herhangi bir yöntem ile kümeleme işleminin gerçekleştirildiği bir çalışmaya rastlanmamış olmakla birlikte, YSA ve DKSBS ile yapılan bir sınıflandırmaya da rastlanmamıştır. Sonuçlara bakıldığı zaman tahminleme doğruluğu bakımından OSB'nin çocuklar için olan verilerinde, sınıflandırma yöntemlerinin kümeleme yöntemlerinden daha başarılı olduğu görülmektedir. Özellikle DKSBS yönteminin tüm verileri doğru sınıflandırarak literatürdeki birçok yöntemden daha yüksek başarı oranına sahip olduğu için tercih edilebilecek yöntemler arasında en iyi yöntemlerden bir tanesi olabileceği sonucuna varılmaktadır. Dolayısı ile bu çalışma ile kullanılan veri seti için farklı yöntemlerin denenmesine imkân tanınarak literatürdeki birçok çalışmadan daha başarılı sonuçlar elde ettiği görülen bir yöntem ile sınıflandırılma yapılmasına imkan tanımaktadır. Bunun yanı sıra elde edilen sonuçların değerlendirilmesi için kullanılan parametreler birçok çalışmadan daha detaylı olarak verilmiştir. Bu durum elde edilen sonuçların daha detaylı yorumlanmasına imkan tanımaktadır.

İleriki çalışmalarda ise veri setinin farklı alt kümeleri kullanılarak veya tüm alt kümeleri aynı anda kullanılarak bir takım analizler gerçekleştirilebilir. Yapılacak çalışmalarda karar destek sistemleri, uzman sistemler, görüntü işleme teknikleri gibi daha farklı teknolojiler ile desteklenen daha geniş kapsamlı veri setlerine uygulanan çalışmalar gerçekleştirilebilir. Verilerin yeniden güncel olarak toplanması için alanında uzman hekim veya hekimler ile birlikte ortak olarak yeni çalışmalar yürütülebilir. Devamında ise elde edilen bulgular ile süreli öğrenmeye devam eden ve normal hayatta da kullanılabilen somut ürünler ortaya konulabilir.

Hakem Değerlendirmesi: Dış bağımsız.

Çıkar Çatışması: Yazarlar çıkar çatışması bildirmemiştir.

Finansal Destek: Yazarlar bu çalışma için finansal destek almadığını beyan etmiştir.

Yazar Katkıları: Çalışma Konsepti/Tasarım- S.Ç., M.Ş.Ç.; Veri Toplama- S.Ç., M.Ş.Ç.; Veri Analizi/Yorumlama- S.Ç., M.Ş.Ç.; Yazı Taslağı- S.Ç., M.Ş.Ç.; İçeriğin Eleştirel İncelemesi- S.Ç., M.Ş.Ç.; Son Onay ve Sorumluluk- S.Ç., M.Ş.Ç.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

Author Contributions: Conception/Design of Study- S.Ç., M.Ş.Ç.; Data Acquisition- S.Ç., M.Ş.Ç.; Data Analysis/Interpretation- S.Ç., M.Ş.Ç.; Drafting Manuscript- S.Ç., M.Ş.Ç.; Critical Revision of Manuscript- S.Ç., M.Ş.Ç.; Final Approval and Accountability- S.Ç., M.Ş.Ç.

Kaynaklar/References

- Akçapınar, G., Altun, A. & Aşkar, P. (2016). Çevrimiçi Öğrenme Ortamındaki Benzer Öğrenci Gruplarının Kümeleme Yöntemi İle Belirlenmesi. *Eğitim Teknolojisi Kuram ve Uygulama*, 6(2), 46-64. DOI: 10.17943/etku.91440.
- Akdemir, Ç. (2016). "Hilenin Veri Madenciliği İle Ortaya Çıkarılması ve Perakende Sektöründe Bir Uygulama", (Yayınlanmamış Doktora Tezi), İstanbul: Marmara Üniversitesi Sosyal Bilimler Enstitüsü.
- Akgöbek, Ö., ve Çakır, F. (2009). Veri Madenciliğinde Bir Uzman Sistem Tasarımı, *Akademik Bilişim Konferansı*, 9, 801-806.
- Akkaya, G. C., Demireli, E., Yakut, Ü. H., & Yakut, H. (2009). İşletmelerde Finansal Başarısızlık Tahminlemesi: Yapay Sinir Ağları Modeli İle İmkân Üzerine Bir Uygulama. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 10(2), 187-216.
- Akyılmaz, O., & Ayan, T. (2010). Esnek Hesaplama Yöntemlerinin Jeodezide Uygulamaları. *İTÜ Dergisi*, 5(1), 261-268.
- Akyol, K., & Karaci, A. (2018). A Study on Autistic Spectrum Disorder For Children Based On Feature Selection And Fuzzy Rule, *International Congress on Engineering and Life Science*, 804–807.
- Aydın, S. (2007). "Veri Madenciliği ve Anadolu Üniversitesi Uzaktan Eğitim Sisteminde Bir Uygulama", (Yayımlanmamış Doktora Tezi). Eskişehir: Anadolu Üniversitesi Sosyal Bilimler Enstitüsü.

- Bholowalia, P., & Kumar, A. (2014). EBK-Means: A Clustering Technique Based On Elbow Method And K-Means In WSN. *International Journal of Computer Applications*, 105(9), 17-24.
- Bradley, P. S., Fayyad, U. M., & Reina, C. (1998). Scaling Clustering Algorithms to Large Databases. *KDD*, 98, 9-15.
- Brossette, S. E., Sprague, A. P., Hardin, J. M., Waites, K. B., Jones, W. T., & Moser, S. A. (1998). Association Rules and Data Mining in Hospital Infection Control And Public Health Surveillance. *Journal of The American Medical Informatics Association*, 5(4), 373-381.
- Budak, E. Ç., & Bozkurt, M. R. (2013). Vertebra Lomber Disklerde Meydana Gelen Bozulmaların Manyetik Rezonans Görüntüleme (Mrg) İle Analizi. *AJIT-e: Online Academic Journal of Information Technology*, 4(11), 125-144.
- Çağlar, M. F., Çetişli, B., & Toprak, İ. B. (2010). Automatic Recognition of Parkinson's Disease From Sustained Phonation Tests Using ANN and Adaptive Neuro-Fuzzy Classifier. *Mühendislik Bilimleri ve Tasarım Dergisi*, 1(2), 59-64.
- Çalışkan, S. K., & Soğukpınar, İ. (2008). Kxknn: K-Means ve K En Yakın Komşu Yöntemleri İle Ağlarda Nüfuz Tespiti. *EMO Yayınları*, 120-24.
- Çelik, S., Bozkurt, Ö. Ç., & Çeşmeli, M. Ş. (2018). İnsan Omurgası Veri Setinin Sinir-Bulanık Sınıflayıcı İle Öznetelik Tespiti ve Sınıflandırılması. *Yönetim Bilişim Sistemleri Dergisi*, 4(1), 39-52.
- Çelik, S. (2020). Determination and Classification of Importance of Attributes Used in Diagnosing Pregnant Women's Birth Method. *Alphanumeric Journal*, 8(2), 261-274.
- Cetişli, B., & Kalkan, H. (2011). Classification of Multispectral Satellite Images By Using Adaptive Neuro-Fuzzy Classifier with Linguistic Hedges. *Signal Processing and Communications Applications Conference*, 19, 50-53.
- Çetişli B. (2006). Öznetelik Seçiminde Dilsel Kuvvetli Sinir Bulanık Sınıflayıcı Kullanımı. *Eskişehir Osmangazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi*, 19(2), 109-130.
- Çınaroğlu, S., & Bulut, H. (2018). New Initialization Approaches for The K-Means And Particle Swarm Optimization Based Clustering Algorithms. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 33(2), 413-423.
- Deb, C., Lee, S. E., & Santamouris, M. (2018). Using Artificial Neural Networks to Assess HVAC Related Energy Saving in Retrofitted Office Buildings. *Solar Energy*, 163, 32-44.
- Demiralay, M., & Çamurcu, A. Y. (2005). Cure, Agnes ve K-Means Algoritmalarındaki Kümeleme Yeteneklerinin Karşılaştırılması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 4(8), 1-18.
- De Campos Souza, P. V., & Guimaraes, A. J. (2018). Using Fuzzy Neural Networks for Improving the Prediction of Children with Autism Through Mobile Devices. *Symposium on Computers and Communications*, 01086-01089.
- Diler, S. (2016). "Veri Madenciliği Süreçleri ve Karar Ağaçları Algoritmaları ile Bir Uygulama", (Yayımlanmamış Yüksek Lisans Tezi). Van: Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü.
- Doğan, B., Buldu, A., Demir, Ö., & Erol, B. (2018). Sigortacılık Sektöründe Müşteri İlişki Yönetimi İçin Kümeleme Analizi. *Karaelmas Fen ve Mühendislik Dergisi*, 8(1), 11-18.
- Feng, Y., & Hamerly, G. (2007). PG-Means: Learning The Number of Clusters in Data. In *Advances in neural information processing systems*, 393-400.
- Gülsöz, T., & Çıkılı, Y. (2018). Otizm Spektrum Bozukluğu Olan Öğrencilere Soğuk İçecek Hazırlama ve Sunma Becerisinin Video Model ile Öğretimin Etkililiği. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 18(1), 210-229.
- Güner, S., Codal, K. S., Geçer, H. S., & Coşkun, E. (2018). Trafik Kaza Desenlerinin Tanımlanmasında K-Means Kümeleme Algoritmasının Kullanılması: Sakarya İli Uygulaması. *İşletme Bilimi Dergisi*, 6(3), 89-105.
- Hamerly, G., & Elkan, C. (2004). Learning The K in K-Means. In *Advances in Neural Information Processing Systems*, 281-288.
- Hong, T. P., & Wu, C. W. (2011). Mining Rules From an Incomplete Dataset with a High Missing Rate. *Expert Systems with Applications*, 38(4), 3931-3936.
- Huang, D., & Chow, T. W. (2005). Efficiently Searching The Important Input Variables Using Bayesian Discriminant. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 52(4), 8785-793.
- Hutchinson, J., Schauer, I., & Seetan, R. (2019). A Comparative Study of Data Mining Techniques Used to Test Predictive Accuracy of Autism Spectrum Disorder Screening Process. In *34th Annual Conference of The Pennsylvania Association of Computer and Information Science Educators*, 70-75.
- Işık, M., & Çamurcu, A. Y. (2007). K-Means, K-Medoids ve Bulanık C-Means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 6(13), 31-45.
- Işık M., & Çamurcu, A. Y. (2011). Document Clustering Using K-Means and Hyperspherical Fuzzy C-Means Algorithms. *Marmara Fen Bilimleri Dergisi*, 22(1), 1-18.
- Jain, A. K. (2010). Data Clustering: 50 Years Beyond K-Means. *Pattern Recognition Letters*, 31(8), 651-666.
- Jayalakshmi, T., & Santhakumaran, A. (2011). Statistical Normalization and Back Propagation for Classification. *International Journal of Computer Theory and Engineering*, 3(1), 1793-8201.
- Kalantar, B., Pradhan, B., Naghibi, S. A., Motevalli, A., & Mansor, S. (2018). Assessment of The Effects of Training Data Selection on The Landslide Susceptibility Mapping: A Comparison Between Support Vector Machine (SVM), Logistic Regression (LR) and Artificial Neural Networks (ANN), Geomatics, *Natural Hazards and Risk*, 9(1), 49-69.
- Kalogeratos, A., & Likas, A. (2012). Dip-Means: an Incremental Clustering Method For Estimating The Number Oo Clusters. In *Advances in Neural Information Processing Systems*, 2393-2401.
- Khaled, A., Sanjay, R., & Vineet, S. (1998). An Efficient K-Means Clustering Algorithm. In *IPPS: 11th International Parallel Processing Symposium*.
- Bilgin, T. T., & Çamurcu, Y. (2005). DbSCAN, Optics ve K-Means Kümeleme Algoritmalarının Uygulamalı Karşılaştırılması. *Politeknik Dergisi*, 8(2), 139-145.

- Kılınc, G. E., & Söğüt, M. Ü. (2018). Mikrobiyotaya Güncel Bir Yaklaşım: Otizm ve Mikrobiyotaya, *Türkiye Klinikleri Journal of Health Sciences*, 3(1), 88-94.
- Koçyiğit, Y., & Korürek, M. (2010). EMG İşaretlerini Dalgacık Dönüşümü ve Bulanık Mantık Sınıflayıcı Kullanarak Sınıflama. *İTÜDERGİSİ/d*, 4(3), 25-31.
- Küçükkoçaoğlu, G., Benli Keskin, Y., & Küçüksözen, C., (2005). Finansal Bilgi Manipülasyonunun Tespitinde Yapay Sinir Ağı Modelinin Kullanımı. *İMKB Dergisi*, 9(36), 1-23.
- Liu, H., Dougherty, E. R., Dy, J. G., Torkkola, K., Tuv, E., Peng, H., ... & Zhao, Z. (2005). Evolving Feature Selection, *IEEE Intelligent systems*, 20(6), 64-76.
- Mackinnon, M. J., & Glick, N. (1999). Applications: Data Mining and Knowledge Discovery in Databases—an Overview. *Australian & New Zealand Journal of Statistics*, 41(3), 255-275.
- MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. In Proceedings of The Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1(14), 281-297.
- Mehmet, D., & Akpınar, E. (2018). Sabit ve Hareketli Hava İle Isıtılan Güneş Kolektörlerinin Isıl Verimliliklerinin Karşılaştırılması ve Yapay Sinir Ağları İle Modellenmesi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 30(1), 41-46.
- Moskov, D. (2016). "Knowledge Extraction From Published Papers in Literature For The Catalytic Methanol Production From Synthesis Gas Using Data Mining Tools", (Yayımlanmamış Yüksek Lisans Tezi). İstanbul: Boğaziçi Üniversitesi. Fen Bilimleri Enstitüsü.
- Muhr, M., & Granitzer, M. (2009). Automatic Cluster Number Selection Using A Split and Merge K-Means Approach, *In 20th International Workshop on Database and Expert Systems Application*, 363-367.
- Nilashi, M., Ibrahim, O., Dalvi, M., Ahmadi, H., & Shahmoradi, L. (2017). Accuracy Improvement For Diabetes Disease Classification: A Case on A Public Medical Dataset. *Fuzzy Information and Engineering*, 9(3), 345-357.
- Oğuzlar, A. (2003). Veri Ön İşleme. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 21, 67-76.
- Olçay Gül, S., & Tekin İftar, E. (2012). Otizm Spektrum Bozukluğu Tanısı Olan Bireyler İçin Sosyal Öykülerin Kullanımı. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 13(2), 1-24.
- Özsoy, İ., & Fırat, M. (2004). Kirişsiz Döşemeli Betonarme Bir Binada Oluşan Yatay Deplasmanın Yapay Sinir Ağları İle Tahmini. *Deü Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 6(1), 51-63.
- Pelleg, D., & Moore, A. W. (2000). X-Means: Extending K-Means with Efficient Estimation of The Number of Clusters, *In Icml*, 1, 727-734.
- Pençe, İ., & Çetişli, B. (2013). El Yazı Karakterlerinin Kapalı Cebirsel Eğrilerle Modellenmesi ve Sınıflandırılması. *Sigma*, 5, 1-7.
- Reddy, S. K., Kodali, S. R., & Gundabathina, J. L. (2012). Classification of Vertebral Column Using Naïve Bayes Technique. *International Journal of Computer Applications*, 58(7), 38-42.
- Sabuncuoğlu, M., Cebeci, S., Rahbar, M. H., & Hessabi, M. (2015). Autism Spectrum Disorder and Attention Deficit Hyperactivity Disorder: Knowledge and Attitude of Family Medicine Residents in Turkey. *Turkish Journal of Family Medicine & Primary Care*, 9(2), 46-53.
- Sariman, G. (2011). Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir Çalışma: K-Means ve K-Medoids Kümeleme Algoritmalarının Karşılaştırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 15(3), 192-202.
- Sebik, N. B., & Bülbül, H. İ. (2018). Veri Madenciliği Modellerinin Akciğer Kanseri Veri Seti Üzerinde Başarılarının İncelenmesi. *TÜBAV Bilim Dergisi*, 11(3), 1-7.
- Sezgin, E., & Çelik, Y. (2013). Veri Madenciliğinde Kayıp Veriler İçin Kullanılan Yöntemlerin Karşılaştırılması. *Akademik Bilişim Konferansı*, 23-25.
- Steinley, D. (2006). K-Means Clustering: A Half-Century Synthesis, *British Journal of Mathematical and Statistical Psychology*, 59(1), 1-34.
- Tsumoto, S., Kimura, T., Iwata, H., & Hirano, S. (2017). Mining Text For Disease Diagnosis. *Procedia Computer Science*, 122, 1133-1140.
- Thabtah, F., Kamalov, F., & Rajab, K. (2018). A New Computational Intelligence Approach to Detect Autistic Features For Autism Screening. *International Journal of Medical Informatics*, 117, 112-124.
- Thabtah, F. (2017). Autism Spectrum Disorder Screening: Machine Learning Adaptation and Dsm-5 Fulfillment. *In Proceedings of the 1st International Conference on Medical and Health Informatics*, 1-6.
- Thabtah, F. (2018). An Accessible and Efficient Autism Screening Method For Behavioural Data and Predictive Analyses. *Health Informatics Journal*, 25(4), 1739-1755.
- Tekin, R., Kaya, Y., & Tağluk, M. E. (2011). K-Means ve Ysa Temelli Hibrit Bir Model İle Epileptik Eeg İşaretlerinin Sınıflandırılması. *Elektrik Elektronik Bilgisayar Sempozyumu*, 277-283.
- Torun, Y. T., & İşeri, E. (2018). Bebeklik Döneminde Önemli Bir Psikopatoloji: Otizm. *Türkiye Klinikleri Journal of Child Psychiatry-Special Topics*, 4(1), 48-55.
- Tsai, C. Y., & Chiu, C. C. (2008). Developing A Feature Weight Self-Adjustment Mechanism For A K-Means Clustering Algorithm. *Computational Statistics & Data Analysis*, 52(10), 4658-4672.
- Tiryaki, A. E., & Kazan, R. (2007). Bulaşık Makinesinin Bulanık Mantık İle Modellenmesi. *Mühendis ve Makine*, 48(565), 3-8.
- Tucker, C. S., Kim, H. M., Barker, D. E., & Zhang, Y. (2010). A Relief Attribute Weighting and X-Means Clustering Methodology For Top-Down Product Family Optimization. *Engineering Optimization*, 42(7), 593-616.
- Uğur, Ç., Sertçelik, M., Üneri, Ö., Dinç, G. Ş., Sekmen, E., & Solmaz, E. (2018). Evaluation of Serum Urotensin-II Levels of Children with ADHD and Autism Spectrum Disorder. *Anadolu Psikiyatri Dergisi*, 19(1), 80-86.
- Uğur, Ç., & Göker, Z. (2018). Bir Çocuk Psikiyatrisi Ünitesine Başvuran Olgularda Otizm Spektrum Bozukluğu: Demografik ve Klinik Bulgular, *Türkiye Çocuk Hastalıkları Dergisi*, 13(3), 177-183.
- Van den Bekerom, B. (2017). Using Machine Learning For Detection of Autism Spectrum Disorder. *In Proc. 20th Student Conf*, 1-7.
- Yavuz, Ü., Ekim, U. & Köklü, M. (2011). Üniversite Öğrencilerin Ortak Zorunlu Derslerdeki Başarılarının K-Means Algoritması İle İncelenmesi. *NWSA: Engineering Sciences*, 6(1), 342-347.
- Yavuz, S., & Deveci, M. (2012). İstatistiksel Normalizasyon Tekniklerinin Yapay Sinir Ağına Performansına Etkisi. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 40, 167-187.