

Otizm Spektrum Bozukluğu Tanısında Öznitelik Seçimi Yoluyla Farklı Sınıflandırıcıların Karşılaştırılması

Sinem KOÇ^{1*}, Onur SEVLİ²

¹ Bilgisayar ve Öğretim Teknolojileri Eğitimi, Eğitim Bilimleri Enstitüsü, Burdur Mehmet Akif Ersoy Üniversitesi, Burdur, Türkiye

² Bilgisayar Mühendisliği Bölümü, Mühendislik-Mimarlık Fakültesi, Burdur Mehmet Akif Ersoy Üniversitesi, Burdur, Türkiye

*1 sinem.kc.07@gmail.com, ² onursevli@mehmetakif.edu.tr

(Geliş/Received: 15/06/2024;

Kabul/Accepted: 17/10/2024)

Öz: Otizm Spektrum Bozukluğu (OSB), sosyal etkileşim ve iletişimde zorluklar, tekrarlayıcı işlemler ve duygusal sorunlar gibi belirgin bir dağılıma sahip karmaşık bir nörogelişimsel durumdur. Bireylerin sosyal etkileşimi, iletişimin gelişmesi ve belirli davranış kalıpları üzerindeki zorluklarla birlikte kendini gösterir. Otizmin genellikle erken çocukluk döneminde başladığı ve bu dönemde belirginlik kazandığı gözlemlenmektedir. Otizmde erken teşhis önemlidir; çünkü erken tanı ile tedavinin erken başlaması mümkündür. OSB sorununun teşhisi için geleneksel yöntemlere ek olarak, günümüzde istikrarlı çıkarımları ile farklı pek çok alanda uygulamaları olan makine öğrenmesi yöntemleri teşhis başarısını artırmak amacıyla kullanılmaktadır. Yöntemler, büyük veri setlerini analiz ederek otizm belirtilerini hızla tanımlamak, erken ve doğru teşhis sağlamak amacıyla bu çalışmada 17 girdi ve 1 hedef olmak üzere toplam 18 öznitelik değişkenden oluşan bir veri seti üzerinde öznitelik seçimi yöntemi ve sınıf dengeleme yöntemleri uygulanarak ardından dört farklı makine öğrenmesi algoritması (K-En Yakın Komşu, Lojistik Regresyon, Naive Bayes, Destek Vektör Makineleri) ile sınıflandırma işlemleri gerçekleştirilmiştir. Sınıflandırma performansı doğruluk, duyarlılık, özgüllük ve F1 skoru gibi metriklerle değerlendirilmiştir. Öznitelik seçimi sonrası, Destek Vektör Makineleri ve Lojistik Regresyon algoritmaları ile %100 doğruluk oranı elde edilirken, K-En Yakın Komşu ve Naive Bayes algoritmaları sırasıyla %94,7 ve %96,7 doğruluk sağlamıştır. Öznitelik seçimi yapılmadığında ise en yüksek doğruluk oranı %96,2 olarak kaydedilmiştir. Sonuçlar, öznitelik seçiminin makine öğrenmesi algoritmalarının sınıflandırma performansını belirgin bir şekilde artırdığını göstermektedir. Bu sonuçlar doğrultusunda OSB tanısında makine öğrenmesi yöntemlerinin uygulanabilirliğini ve doğruluğunu ortaya koymakta olup, teşhis sürecini iyileştirmek adına önemli bir katkı sağlamaktadır.

Anahtar kelimeler: Makine Öğrenmesi, Otizm Spektrum Bozukluğu, Otizm Teşhisi, Öznitelik Seçimi.

Comparison of Different Classifiers Through Feature Selection in The Diagnosis of Autism Spectrum Disorder

Abstract: Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition with a marked distribution of difficulties in social interaction and communication, repetitive processing and emotional problems. It is manifested by difficulties with individuals' social interaction, the development of communication and difficulties with certain patterns of behavior. It is observed that autism usually starts in early childhood and gains prominence in this period. Early diagnosis is important in autism because early diagnosis allows treatment to start early. In addition to traditional methods for diagnosing ASD, machine learning methods, which have applications in many different fields with their stable inferences, are now used to increase diagnostic success. In this study, SelectKBest feature selection method was applied using a dataset consisting of 18 feature variables (17 input and 1 target) and classification processes were performed with four different machine learning algorithms (K-Nearest Neighbor, Logistic Regression, Naive Bayes, Support Vector Machines). Classification performance was evaluated with metrics such as accuracy, sensitivity, specificity and F1 score. After feature selection, Support Vector Machines and Logistic Regression algorithms achieved 100% accuracy, while K-Nearest Neighbor and Naive Bayes algorithms achieved 94.7% and 96.7% accuracy, respectively. When no feature selection was performed, the highest accuracy rate was recorded as 96.2%. The results show that feature selection significantly improves the classification performance of machine learning algorithms. These results demonstrate the applicability and accuracy of machine learning methods in the diagnosis of ASD and provide an important contribution to improve the diagnostic process.

Key words: Machine Learning, Autism Spectrum Disorder, Autism Diagnosis, Feature selection.

* Sorumlu yazar: sinem.kc.07@gmail.com. Yazarların ORCID Numarası: ¹ 0009-0002-8451-6452, ² 0000-0002-8933-8395

1. Giriş

Otizm Spektrum Bozukluğu (OSB), nörogelişimsel bir durum olup, genellikle çocukluk döneminde başlayan ve ömür boyu süren karmaşık bir bozukluktur. Bu bozukluk, bireylerin sosyal etkileşim becerilerinde, iletişim yeteneklerinde ve davranışlarında belirgin farklılıkların ortaya çıktığı geniş bir spektrumu kapsar. OSB'nin belirtileri kişiden kişiye önemli ölçüde değişkenlik gösterir ve bu durum, genetik ve çevresel faktörlerin etkileşimiyle şekillenen karmaşık bir etiyolojiye işaret eder. Genellikle tipik gelişim gösteren bireylerle kıyaslandığında, OSB tanısı alan bireyler tekrarlayıcı davranışlar, sınırlı ilgi alanları ve sözel olmayan iletişim becerilerinde zorluklar yaşarlar. OSB'nin nedenleri konusundaki araştırmalar, genetik yatkınlığın yanı sıra çevresel etkenlerin de etkili olduğunu göstermektedir. Bu bağlamda, OSB karmaşık ve çeşitliliği yüksek bir bozukluktur [1]. Otizmin belirtileri genellikle çocuğun ilk yaşlarında ortaya çıkar ve genellikle belirgin bir şekilde 18 ila 24 ay arasında görülür. Ancak, bu belirtiler bazı durumlarda doğumdan sonraki ilk yıl içinde de fark edilebilir. Birçok otizmlili çocuk, sosyal etkileşim ve iletişim becerilerindeki zorluklar, tekrarlayıcı davranışlar ve sınırlı ilgi alanları gibi belirtilerle dikkat çeker. Ancak, dikkat çekici olan, bazı çocukların doğumdan sonra normal bir gelişim gösteriyor gibi görünmeleridir. Bu çocuklarda, genellikle 18 ila 24 aylık dönemde, bir gerileme yaşanır ve otizme özgü belirtiler belirginleşmeye başlar. Erken bebeklik otizmi, ilk defa 1943 yılında Leo Kanner tarafından tanımlanmıştır. Kanner'ın gözlemlerine dayanarak yapılan bu tanımlama, otistik çocukların doğumdan itibaren diğer insanlarla olan ilişki becerilerinin zayıf olduğunu, belirgin ekolaliye sahip olduklarını, zamirleri yanlış kullandıklarını, davranışlarında tekrarlayıcı ve amaçsız etkinliklere yönelik bir eğilim gösterdiklerini, cansız objelere karşı belirgin bir ilgi sergilediklerini ve değişikliklere karşı huzursuz olduklarını vurgulamaktadır [1]. Bu tanımlama, erken bebeklik otizminin temel özelliklerini açıklamakla birlikte, günümüzdeki bilimsel literatür ve klinik gözlemler ışığında genişlemiş ve derinleşmiştir [2]. OSB terimi, farklı bireyler arasındaki çeşitliliği ve belirtilerin şiddet derecesini daha iyi ifade ederek Kanner'ın orijinal tanımına evrilmiştir. Bu bağlamda, erken bebeklik otizmi üzerine yapılan çağdaş araştırmalar, nörolojik, genetik ve çevresel faktörlerin etkileşimini daha detaylı bir şekilde incelemekte ve bireysel farklılıkları dikkate almaktadır [3]. OSB'de erken teşhis son derece önemlidir. Erken müdahalenin genel hedefi, bireyin gelişimsel, iletişimsel ve sosyal becerilerini olumlu yönde etkilemektir. Erken müdahale, genellikle çocukluk döneminde başlar, zira bu dönemde beyin plastisitesi yüksektir ve bireyin öğrenme yetenekleri daha fazla esnekler. Erken müdahale, bireyin gelişimini destekleme, dil ve iletişim becerilerini artırma ve çevresiyle etkileşim kurma yeteneğini iyileştirmeye yönelik stratejileri içerir [4]. Otizm tanısı almış bireylerin aileleri, genellikle bir dizi benzersiz zorlukla karşı karşıyadır. Bu durum, aile üyelerinin günlük yaşam, eğitim, sosyal etkileşim ve gelecek planlaması konularında çeşitli önemli kararlar almasını gerektirir [5]. OSB olan bir çocuğa sahip olmak, aile dinamiklerini etkileyebilir ve çeşitli duygusal, finansal ve sosyal zorluklara neden olabilir [6]. Bireylerin eğitimine mümkün olan en erken aşamada başlanması, bireyin gelişimi açısından kritik bir öneme sahiptir. Ancak, OSB'nin teşhisi karmaşık ve zaman alıcı bir süreç olduğundan, tedaviye başlama noktasında gecikmeler yaşanabilmektedir. Hastalığın teşhisi sürecindeki belirsizlikleri ve gecikmeleri azaltmak amacıyla, günümüzde alternatif karar destek yöntemlere başvurulmaktadır.

Yapay zekâ tekniklerinin günümüzde pek çok alanda başarılı uygulamaları mevcuttur. Yapay zekânın bir alt disiplini olan makine öğrenmesi mevcut veriler üzerinde, bazen keşfedilmesi insan eliyle son derece zor olan örüntüleri bularak gelecek durumlar hakkında istikrarlı tahminlerde bulunmayı sağlayan teknikleri içerir [7]. Bu durum, otizmin erken teşhisine yönelik çabaları güçlendirmeyi amaçlayan bir yaklaşım olarak da ön plana çıkmaktadır. Makine öğrenmesi teknikleri, belirli davranış örüntülerini analiz ederek otizm belirtilerini daha hızlı ve hassas bir şekilde tanımlama potansiyeline sahiptir. Bu klinik uzmanların, özellikle gelişimsel belirtilerin ortaya çıktığı erken çocukluk dönemlerinde, doğru teşhisi koyabilmeleri için önemli bir destek sistemi rolü üstlenebilir. Bu teknikler, büyük veri setlerini analiz ederek otizmin belirli özelliklerini daha iyi anlamaya yardımcı olabilmektedir. Özellikle, semptomların çeşitliliği ve bireyden bireye farklılık göstermesi nedeniyle, makine öğrenmesi, geniş veri setlerini işleyerek, öznitelik seçimi yaparak daha kesin ve bireyselleştirilmiş teşhisler sunma potansiyeline sahiptir. Bu yöntem, teşhis sürecindeki gecikmeleri azaltarak ve daha kesin sonuçlar elde ederek OSB'li bireylere erken ve etkili müdahale imkânı sunar. Bu çalışmada, OSB'li ve OSB'siz bireylerin ayırt edilmesi amacıyla SelectBest öznitelik seçimi yöntemi kullanılmıştır. OSB teşhisi sürecindeki belirsizlikleri azaltarak ve daha hızlı sonuçlar üreterek klinik uzmanların kararlarını destekleyici bir sistem sunmaktadır.

Bu çalışmanın amaçları şunları içerir:

- Makine öğrenmesi teknikleri kullanılarak OSB'nin erken ve etkin teşhisi: Otizm spektrum bozukluğunun erken teşhisinde yapay zekâ ve makine öğrenmesi tekniklerinin etkinliğini inceleyerek bu alandaki mevcut literatüre katkıda bulunmaktadır.
- Öznitelik seçimi ve sınıf dengeleme ile performanslı ve doğruluğu yüksek teşhis: Mevcut veri seti üzerinde öznitelik seçimi yapılarak ve sınıf dengesi sağlanarak sınıflandırma performansı artırılmıştır.

- Teşhis Süreçlerindeki Gecikmelerin Azaltılması: Yapay zekâ tabanlı karar destek sistemlerinin, otizm teşhisindeki süreçleri hızlandırarak, klinik gecikmeleri nasıl azaltabileceği tartışılmıştır.

Literatürde OSB teşhisine yönelik yapılan çalışmalar, farklı makine öğrenmesi uygulamalarının kullanımı geniş bir şekilde yer verilmektedir.

Çelik ve Çeşmeli'nin [8] çalışması, OSB teşhisinde veri madenciliği yöntemlerinin uygulanabilirliğini araştırmaktadır. 292 OSB'li çocuktan elde edilen gerçek veriler ile 20 öznitelik ve 1 çıkış özelliği kullanılarak, yapay sinir ağları ve dilsel kuvvetli sinir-bulanık sınıflandırıcılarla ilişkilendirilmiştir. K-means ve X-means kümelenmeleri ile veriler üzerinde analiz gerçekleştirilmiş, OSB teşhis özellikleri katkı sunabileceği vurgulanmıştır. Demirhan [9] OSB'nin ergenlerde farklı makine öğrenme teknikleri ile incelenmesini ve karşılaştırılmasını gerçekleştirmiştir. DVM, KNN ve RO çözümleri kullanılarak yapılan geliştirmeler, RO yöntemi %100 doğruluk oranı ile öne çıkan ve bu bulgular, RO'nun klinik uygulamalarında OSB teşhisinde güvenilir bir araç olarak kullanılabilirliğini göstermektedir. Metlek ve Kayaalp [10] YSA ve dilsel kuvvetli sinir-bulanık sınıflandırıcıyı kullanarak %100'de başarı oranı göstermiştir. Bununla birlikte YSA'nın test verisinde %98.85 başarı oranı ile daha üstün bir görünüm yeteneği sergilediği vurgulanmıştır.

Thabtah ve Peebles [11] Lojistik Regresyon (LR) ile %97,94 ve Naive Bayes (NB) ile %92,80 başarı oranlarına ulaşmıştır. Akyol ve Karaci [12] ise genel kurallı lojistik regresyon raporları ile %92,00 ve %97,33 başarı oranları sağlamıştır. Saihi ve Alshraideh [14] çalışmalarında, C4.5, RO ve Sinir Ağı modellerini karşılaştırmış ve Sinir Ağı'nın en yüksek doğruluk performansına sahiptir. Şuvo vd [15] RO sınıflandırıcısını 10 davranışsal ve 10 bireysel özellik içeren veri seti üzerinde inceleyerek %0,96 genel doğruluk elde etmiştir. Bidwe ve arkadaşları [16] Zenodo'dan alınan yüzde verileri, veri ayarları Üzerinde Transfer Öğrenme (TL) yöntemlerini kullanarak, VGG16, VGG19, InceptionV3, AlexNet ve ResNet50 modelleri arasında InceptionV3'ün %87,99 doğruluk oranı ile en yüksek performansı gösterdiğini saptamışlardır.

Raj ve Masood [17] NB, DVM, LR, KNN, Sinir Ağı (NN) ve Evrimsel Sinir Ağı (CNN) modelleriyle çocukluk, ergenlik ve yetişkin OSB öngörüsü yapmayı amaçlamış ve CNN'in %99,53, %98, 30 ve %96,88 doğruluk oranlarıyla üstün performans sergilediğini belirtmişlerdir. Alaika ve Alamsyah [18] DVM ve CFS (Korelasyon Tabanlı Öznitelik Seçimi) kullanılmıştır. UCI makine öğrenimi deposundan alınan Çocuklara Yönelik Otistik Spektrum Bozukluğu Taraması Veri Seti üzerinde yapılan karşılaştırmada, DVM'nin %94,91 doğruluk oranı sağladığı tespit edilmiştir. DVM algoritması, CFS ile birleştirildiğinde ise doğruluk oranı %96,61'e yükselerek, seçilen 17 öz niteliğin kullanılmasıyla %1,7'lik bir iyileşme göstermektedir. Saran ve Pirouz [19] OSB üzerine odaklanarak EEG sinyallerinden duyguların çıkarılması ve KNN, RO, DVM ve LR kullanılarak, OSB özelliklerinin tahmin edilmesini içeriyor. Yeni yürümeye başlayan çocuklar ve yetişkinlerde OSB verilerine dayanarak geliştirilen model %73 doğruluk oranına ulaştığı göstermektedir. Misman vd [20] OSB tanısında performansı artırmak amacıyla DNN mimarisini kullanarak yetişkin OSB tarama verilerini analiz etmektedir. İki farklı veri seti üzerinde yapılan analizlerde DNN modelinin %99,40 ve %96,08 doğruluk oranları elde ettiği gözlemlenmiştir. Karşılaştırmalı olarak, DVM modeli %95,24 ve %95,08 doğruluk oranlarına ulaşmıştır.

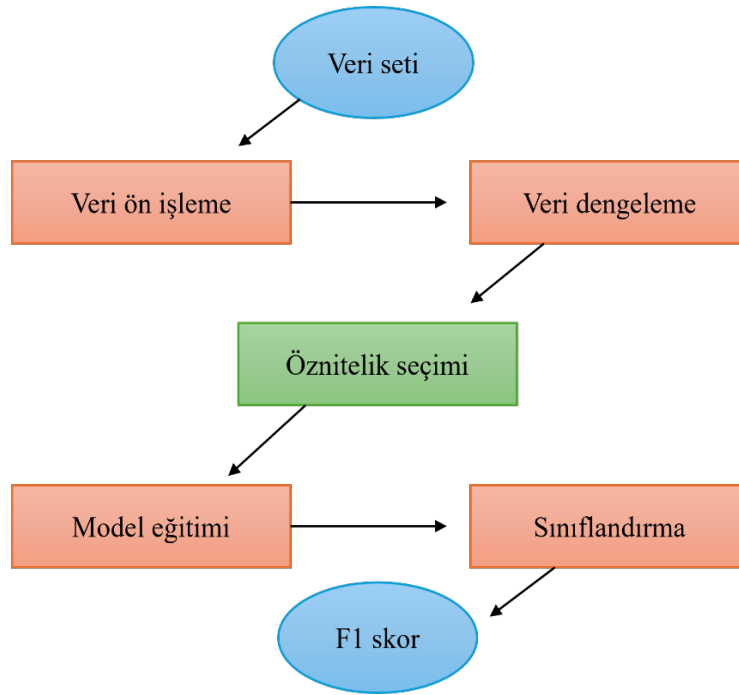
Khatun ve ekibi [21] 704 klinik OSB verisi üzerinde dört denetimli öğrenme tekniğini kullanarak DVM ile %85 doğruluk oranı sağlamış, en yüksek performansla ulaşmıştır. Santos [22] KNN, RO ve DVM karşılaştırmalarını yaparak, %79,35 F1 skoru ile Çok Katmanlı Perceptron'un (MLP) OSB teşhisinde etkili model olduğunu belirtmiştir. Mokni ve Haoues [23] OSB Test veri tabanından elde edilen çocuk, ergen ve yetişkin veri kümeleri üzerinde yapılmıştır. Elde edilen sonuçlar, derin sınıflandırıcı modelinin diğer makine öğrenimi tekniklerinden daha yüksek doğruluk oranlarına sahip olduğunu göstermektedir. Dedgaonkar ve arkadaşları [24] duyuşal işleme bozukluğunun tespit edilmesi ve duygu tanıma amaçlı derin öğrenme teknikleri kullanılmış ve %100'ünün ortaya çıktığı görülmüştür. Uddin [25] RO, Çok Terimli NB, Bernoulli NB, DVM ve GNB yükseltmelerini kullanarak çocuk ve yetişkin veri kümelerinde %100 doğrulukta ulaşmış, özellikle RO'nun en yüksek performansı gösterdiğini belirtmiştir. Praveena ve arkadaşları [26] OSB'nin erken tahmin için farklı özellik dönüştürme tekniklerini incelemiş ve Adaboost'un %100 doğrulukla en etkili sonuçların elde edilmesini sağlamıştır. Özellikle Log, Power-Box Cox ve Yeo-Johnson ayar yöntemleri veri setlerinde başarılı olmuştur. Pottam ve arkadaşları [27] OSB tahmini için Karar Ağacı sınıflandırıcısının %97,47 test performansı ile öne çıktığını belirtmiştir. Günay ve Alkan [28] dört farklı hareket tipine ait elektromiyografi (EMG) sinyallerini gösterirken Destek Vektör Makineleri (DVM) kullanarak, spektral ön işleme tekniklerinin işleyişinin artırdığını ve %94,25 başarı oranının elde edildiğini bildirmiştir. Son olarak, Balcı ve Alkan [29] HPV kaynaklı sigillerin tedavisinde en uygun yöntemle belirlenmesi amacıyla kriyoterapi ve immünoterapi yöntemlerinin yöntemini değerlendirmiştir. HPV'nin biyolojik verim sisteminin atlatılmasına neden olduğu ve hızlı gerekli tedavi olduğu vurgulanmıştır.

Literatür incelemeleri karşılaştırıldığında, çalışmanın hem yüksek doğruluk oranı hem de daha az sayıda öznitelik kullanarak iki farklı modelle üstün başarı sergilemiştir. Özellikle öznitelik seçiminin optimize edilmesi ve bu süreç sonunda elde edilen yüksek doğruluk oranı, çalışmanın literatürdeki diğer araştırmalara kıyasla öne çıkan güçlü yönlerinden biri olarak değerlendirilebilir. Bu durum, çalışmanın verimliliğini artıran önemli bir

avantaj haline getirmiştir. Bu çalışmanın literatüre katkısı, kullanılan yöntemin, daha önce benzer başarımları elde eden yöntemlerle karşılaştırıldığında, belirli avantajlar sunmasıyla öne çıkmaktadır. Öncelikle, önerilen model, çözüm hızını artırırken doğruluğu da optimize etmektedir. Daha önceki çalışmalarda kullanılan yöntemler arasında sıklıkla karar ağaçları, rastgele ormanlar veya destek vektör makineleri yer almakta, ancak bu çalışmada hibrit bir yaklaşım benimsenerek farklı öğrenme algoritmalarının bir arada kullanıldığı görülmektedir. Bu entegrasyon hem işlem süresini kısaltmış hem de doğruluk oranlarında gözle görülür bir artış sağlamıştır. Literatüre yapılan yenilikçi katkı, özellikle makine öğrenmesi modellerinin birleştirilmesiyle elde edilen başarımların artışında yatmaktadır. Bu çalışmanın avantajı, diğer yaklaşımların aksine hem sınıflandırma başarısını artırmak hem de karmaşık veri setleriyle daha verimli başa çıkmak olarak ifade edilebilir.

2. Materyal ve Yöntem

Bu çalışmada kullanılan veri seti, 17 girdi ve 1 hedef değişken olmak üzere toplamda 18 öznitelikten oluşmaktadır. Veri setindeki sınıf dengesizliklerini gidermek amacıyla azınlık sınıf örneklerinin arttırımına dayalı SMOTE veri dengeleme tekniği uygulanmıştır. İstatistiksel testlere dayalı SelectKBest yöntemi kullanılarak, veri setindeki düşük öneme sahip öznitelikler elimine edilerek sınıflandırma performansının artırılması hedeflenmiştir. Veri seti, dört farklı sınıflandırma algoritması ile analiz edilmiştir. Bu algoritmaların her biri, verinin farklı özelliklerini modelleyebilmek adına ayrı ayrı değerlendirilmiş ve karşılaştırılmıştır. Çalışmada uygulanan yöntemin özet akış diyagramı Şekil 1’de verilmiştir.



Şekil 1. Akış diyagramı.

2. 1. Veri Seti

Bu çalışmada OSB'nin erken teşhis ve sınıflandırmasına yönelik olarak, Manukau Enstitüsü'nden araştırmacı Dr Fadi Fayed Thabtah tarafından [30] derlenen ve kamuya açık olarak paylaşılan *Autism Screening Data for Toddlers* [31] bir veri seti kullanılmıştır. Veri seti erken bebeklik dönemindeki otizm tespitine yönelik olarak 12-36 ay arasındaki bebeklerdeki ölçüm sonuçlarını içermektedir. Veri seti içerisinde, her biri 18 farklı öznitelikten oluşan toplam 1054 kayıt bulunmaktadır. Veri seti içerisinde OSB tespitine yönelik olarak uygulanan 10 soruluk testin cevapları ve bunun yanında bireyin çeşitli demografik verileri yer almaktadır. Veri seti içerisinde 17 girdi ve 1 hedef olmak üzere toplam 18 öznitelik bulunmaktadır. Özniteliklerin ayrıntıları Tablo 1’de verilmiştir.

Tablo 1. Veri seti öznitelikler.

Sıra	Öznitelik Adı	Öznitelik Türü	Açıklama
1	A1: Soru	(0,1 şeklindeki var yok verileri tür açısından kategoriktir)	Çocuğunuz adını söylediğinizde size bakıyor mu?
2	A2: Soru	(0,1 şeklindeki var yok verileri tür açısından kategoriktir)	Çocuğunuz ile göz teması kurmak sizin için ne kadar kolay?
3	A3: Soru	(0,1 şeklindeki var yok verileri tür açısından kategoriktir)	Çocuğunuz bir şey istediğini belirtmek için işaret ediyor mu? (örneğin, bir oyuncuğa ulaşamadığında)
4	A4: Soru	(0,1 şeklindeki var yok verileri tür açısından kategoriktir)	Çocuğunuz sizinle ilgisini paylaşmak için işaret ediyor mu? (örneğin, bir şeyi işaret etmek)
5	A5: Soru	(0,1 şeklindeki var yok verileri tür açısından kategoriktir)	Çocuğunuz rol yapıyor mu? (örneğin, oyuncak bebeklere bakmak, oyuncak telefonla konuşmak)
6	A6: Soru	(0,1 şeklindeki var yok verileri tür açısından kategoriktir)	Çocuğunuz nereye baktığınızı takip ediyor mu?
7	A7: Soru	(0,1 şeklindeki var yok verileri tür açısından kategoriktir)	Ailenizden bir birey veya siz gözlemlenebilir bir şekilde üzgün olduğunuzda, çocuğunuzun bu duygu durumunu fark edip aşağıdaki belirtileri gösteriyor mu? (saçlarını okşamak veya sarılmak)
8	A8: Soru	(0,1 şeklindeki var yok verileri tür açısından kategoriktir)	Çocuğunuzun ilk kelimelerini söylemi tanımlarsınız?
9	A9: Soru	(0,1 şeklindeki var yok verileri tür açısından kategoriktir)	Çocuğunuz basit jestler kullanıyor mu? (örneğin, el sallamak)
10	A10: Soru	(0,1 şeklindeki var yok verileri tür açısından kategoriktir)	Çocuğunuz görünürde bir amacı olmadan boş boş bakıyor mu?
11	Age (Yaş)	(sayısal sürekli)	Yaş
12	Score by Q-chat-10 (Q-chat-10'a göre puan)	Tam sayı	Teste cevaplanan sorulara göre elde edilen skor
13	Sex (Cinsiyet)	karakter	Kadın, Erkek
14	Ethnicity (Etnik köken)	Evet hayır	Bireyin etnik kökeni
15	Born with jaundice (Sarılık durumu)	Evet hayır	Birey sarılık geçirmiş mi?
16	Family member with ASD history (OSB geçmişi olan aile üyesi)	Evet hayır	Ailede OSB öyküsü var mı?
17	Who is completing the test (Testi kim tamamlıyor)	string	Aile, sağlık personeli vb..
18	Class variable (Sınıf değişkeni)	string	OSB durumu

OSB'de önemli değişkenlerden biri olan cinsiyetin veri seti içerisindeki örnek dağılımı Şekil 2'de sunulmuştur. Bu grafik, çalışmanın başlangıcında kullanılan veri setindeki cinsiyet dağılımını göstermektedir.



Şekil 2. Cinsiyete göre dağılım grafiği.

Veri setinde yer alan kayıtlardan bireylerin OSB durumunu gösteren hedef değişkene göre dağılımları Şekil 3'te verilmiştir.



Şekil 3. Veri setinde OSB'li olan ve olmayan bireylerin dağılımı.

2.2. Sınıflandırmada Kullanılan Algoritmalar

Makine öğrenmesi, algoritmaların istatistiksel ve matematiksel yöntemleri kullanarak, veri setlerinden öğrenme yeteneğine sahip olduğu bir bilgi işleme dalıdır. Bu algoritmalar, verilerden örüntüler çıkarmak ve gelecekteki bilinmeyen verileri tahmin etmek için kullanılır [32]. Bu çalışmada, sınıflandırma alanında literatürde de yaygın olarak kullanılan dört farklı makine öğrenimi algoritması kullanılmıştır. Bunlar; K-En Yakın Komşu

(K-Nearest Neighbor / KNN), Destek Vektör Makinesi (DVM), Lojistik Regresyon (LR) ve Naive Bayes (NB)'dir. KNN algoritması, sınıflandırma görevlerinde kullanılan bir temel algoritmadır [33]. Eğitim aşamasında, veri noktalarının konumları depolanır. Sınıflandırma sırasında, yeni bir veri noktasının etrafındaki en yakın K komşu belirlenir. Bu komşuların sınıfları üzerinden, yeni veri noktasının sınıfı tahmin edilir.

DVM, iki sınıf arasındaki ayrımı en üst düzeye çıkarmayı hedefleyen bir lineer sınıflandırma veya regresyon modelidir. Özellikle, yüksek boyutlu veri setlerinde etkilidir. Hiper düzlemi bulma esasına dayanır ve bu hiper düzlem, destek vektörleri tarafından belirlenir [34].

LR, bir veri noktasının belirli bir sınıfa ait olma olasılığını tahmin etmek için kullanılan bir lineer modeldir. İkili sınıflandırma görevlerinde yaygın olarak kullanılır. Model, özniteliklerin ağırlıkları ile lineer bir kombinasyonunu alır ve bu değeri bir lojistik fonksiyon içinde kullanarak sınıflandırma gerçekleştirir [35].

NB, Bayes Teoremi'ne dayalı bir olasılık temelli sınıflandırma algoritmasıdır. Öznitelikler arasındaki bağımsızlık varsayımını kullanır. Sınıflandırma sırasında, NB Teoremi ile veri noktasının ait olma olasılığı hesaplanır ve bu olasılıklar üzerinden en yüksek olasılığa sahip sınıf tahmin edilir [36].

2.3. Öznitelik Seçimi

Öznitelik seçimi, makine öğrenmesinde veri boyutunu optimize etmek, sınıflandırma doğruluğunu artırmak ve hesaplama maliyetini minimize etmek amacıyla kullanılan bir stratejidir. Bu yaklaşım, ilgili öznitelikleri seçerek modelin karmaşıklığını azaltır ve dolayısıyla daha etkili bir performans elde etmeyi hedefler. Bu süreç, öznitelikler arasında anlamsız veya gereksiz olanları elemek suretiyle modelin genel performansını artırır [37].

Bu sayede hem işleme hızı artar hem de daha az veriyle daha güvenilir sonuçlar elde etmek mümkün olur. Öznitelik seçimi, veri madenciliği ve makine öğrenimi alanlarında önemli bir aşama olarak kabul edilen bir yöntemdir. Bu yöntem, bir veri kümesinde bulunan özniteliklerin (features) sayısını azaltmayı veya belirli öznitelikleri seçmeyi amaçlar. Genellikle, bir veri setinde bulunan tüm özellikleri kullanmak, modelin karmaşıklığını artırabilir, eğitim sürecini zorlaştırabilir ve genelleme yeteneğini olumsuz yönde etkileyebilir. Bu nedenle, öznitelik seçimi, veri setindeki en önemli ve bilgi taşıyan öznitelikleri belirleyerek, modelin performansını artırmak ve gereksiz karmaşıklığı azaltmak için kullanılır. Öznitelik seçiminin temel amacı, modelin doğruluğunu artırırken aynı zamanda işlem maliyetini düşürmektir. Büyük veri setlerinde bulunan gereksiz veya birbirleriyle düşük korelasyona sahip özelliklerin elemesi, modelin daha hızlı eğitilmesine ve daha etkili çalışmasına olanak tanır. Ayrıca, öznitelik seçimi, modelin anlaşılabilirliğini artırarak, karar süreçlerini daha şeffaf hale getirir [38].

Bu çalışmada öznitelik seçimi yöntemleri arasında "SelectKBest" yöntemi kullanılmıştır. Bu yöntem, scikit-learn kütüphanesinde bulunan bir fonksiyondur ve öznitelik seçiminde temel bir araç olarak kullanılır. SelectKBest yöntemi, veri kümesindeki öznitelikleri belirli bir kriterle değerlendirerek en iyi k özniteliği seçmeyi sağlar. Özniteliklerin değerlendirilmesi için çeşitli istatistiksel metrikler kullanılır. Bu aşamada ise "chi2" istatistiği tercih edilmiştir. Bu istatistik, özellikler arasındaki ilişkiyi ölçer ve en bilgi verici öznitelikleri seçmek için kullanılır. SelectKBest yöntemi, veri kümesindeki öznitelikleri değerlendirme ve seçme konusunda etkili bir yöntemdir. Bu yöntem özellikle sınıflandırma problemlerinde kullanılmak üzere tasarlanmıştır ve modelin daha doğru ve verimli olmasını sağlar. Seçilen öznitelikler, modelin performansını artırmak ve gereksiz öznitelikleri elemek amacıyla kullanılır. Bu seçim işlemi, modelin doğruluğunu artırmak ve analizde daha etkili sonuçlar elde etmek için gereklidir. SelectKBest yöntemiyle öznitelik seçimi yapılırken, chi2 istatistiğini kullanarak özniteliklerin önemi değerlendirilir. Bu istatistik öznitelikler arasındaki ilişkiyi ölçerek en bilgi verici olanların seçilmesini sağlar. Bu süreç, modelin genelleme yeteneğini iyileştirmek ve gereksiz bilgi karmaşasını azaltmak amacıyla önemli bir adımdır.

Öznitelik seçimi sürecinde "SelectKBest" gibi otomatik yöntemler kullanılarak belirlenen özniteliklerin ardından manuel olarak gözden geçirilmesi, modelin doğruluğunu artırmak için önemli bir adımdır. Bu süreç, özniteliklerin daha derinlemesine incelenmesini ve gerektiğinde ayarlanmasını sağlayarak modelin daha güçlü ve güvenilir olmasını destekler.

2.4. Smote Yöntemi

SMOTE, 2002 yılında Chawla ve arkadaşları tarafından önerilen bir fazla örnekleme (oversampling) tekniğidir ve dengesiz veri setlerindeki azınlık sınıfının örnek sayısını artırmak için sentetik veri örnekleri oluşturur [39]. SMOTE'un temel çalışma prensibi, azınlık sınıfındaki örnekler arasında doğrusal interpolasyon yoluyla yeni örnekler üretmektir. Bu yöntem, azınlık sınıfındaki örnekleri çoğaltmak yerine, mevcut veriler arasında yeni sentetik örnekler oluşturarak veri setindeki dengesizliği giderir. Bu çalışmada, sınıf dengesizliğini gidermek ve modelin her iki sınıftan da eşit derecede öğrenmesini sağlamak için SMOTE yöntemi kullanılmıştır. Öncelikle,

veri seti eğitim ve test setlerine ayrılmıştır. Daha sonra, SMOTE uygulanarak eğitim setinde azınlık sınıfı olan OSB tanısı olmayan bireylerin sayısı artırılmıştır. Böylece, modelin azınlık sınıfını göz ardı etme olasılığı azaltılmış ve sınıflar arasında denge sağlanmıştır. SMOTE, sınıf dağılımında denge sağlamak amacıyla azınlık sınıfına yakın olan komşu noktalar arasında yeni veriler oluşturarak çoğaltma yapar. Bu yöntem, özellikle büyük veri setlerinde modelin daha genel bir öğrenme yapmasını sağlayarak aşırı uyum (overfitting) riskini azaltabilir [40]. SMOTE uygulanmadan önce, OSB tanısı olan bireylerin oranı yaklaşık %69 iken, OSB tanısı olmayan bireylerin oranı %31'dir. SMOTE uygulamasıyla bu oran dengelenmiş ve eğitim setindeki sınıf dağılımı eşitlenmiştir. Bu sayede model, her iki sınıfa da eşit önem vererek daha dengeli bir sınıflandırma yapabilecektir. SMOTE ile yapılan oversampling, dengesiz veri setlerinde önemli bir rol oynar ve bu çalışmada veri dengesizliğinin önüne geçmek için etkin bir şekilde kullanılmıştır. Bu teknik sayesinde, model hem OSB tanısı olan hem de olmayan bireyleri daha iyi öğrenerek, sınıflandırma performansını artıracaktır. Literatürde yaygın olarak kullanılan bir oversampling yöntemidir ve dengesiz veri setlerindeki azınlık sınıfına ait örneklerin sayısını artırarak sınıf dengesini sağlamayı hedefler.

2.5. Çapraz Doğrulama

Modellerin etkin bir şekilde değerlendirilmesi için, eğitim seti üzerinde katmanlı 5 katlı çapraz doğrulama yöntemi kullanıldı. Bu yöntem, eğitim setini beş farklı alt kümeye (fold) ayırarak her katmanı hem eğitim hem de doğrulama için kullanır. Her bir katmanın hedef sınıflarının dağılımının tüm veri setini temsil etmesi, çapraz doğrulama sürecinde sınıf dengesizliği ile ilgili endişelerin belirlenmesine katkı sağlar.

Bu yöntem, veri setinin beş alt küme (fold) olarak bölünmesini sağlayarak modelin her bir alt küme üzerinde ayrı ayrı eğitilip değerlendirilmesini mümkün kılmıştır. Modelin güvenilirliğini artırmak ve aşırı uyum (overfitting) problemlerini önlemek amacıyla kullanılır. Bu bütünlüklü yaklaşım, özellikle büyük veri setleri üzerinde etkili, verimli ve genelleme yeteneği yüksek modellerin oluşturulmasına olanak tanımaktadır.

3. Deneysel Çalışma ve Bulgular

Araştırmada sınıflandırma öncesinde veri seti üzerinde ön işleme gerçekleştirilmiş, ardından öznitelik seçimi yapılarak farklı sınıflandırıcılar ile tahminlemeler gerçekleştirilmiştir. Gerçekleştirilen sınıflandırmalar için LR, NB, DVM ve KNN algoritmaları kullanılmıştır.

Veri setinde yer alan 17 adet girdi özneliğinin önce tamamı kullanılarak ve ardından öznitelik seçimi uygulanarak gerçekleştirilen toplam 4 farklı algoritma ile gerçekleştirilen sınıflandırmaların sonuçları çeşitli metrikler açısından değerlendirilmiştir. Sınıflandırmada kullanılan girdi özneliklerin tamamı Tablo 2'de belirtilmiştir.

Tablo 2. Sınıflandırmada kullanılan girdi öznelikleri.

Öznitelik No	Öznitelik Adı	Öznitelik No	Öznitelik Adı
1	A1: SORU	10	A10: SORU
2	A2: SORU	11	A11: SORU
3	A3: SORU	12	A12: SORU
4	A4: SORU	13	A13: SORU
5	A5: SORU	14	A14: SORU
6	A6: SORU	15	A15: SORU
7	A7: SORU	16	A16: SORU
8	A8: SORU	17	A17: SORU
9	A9: SORU	18	

Bu çalışmada 17 adet giriş özneliğinin OSB veri seti üzerinde sınıflandırmalar gerçekleştirilmiştir. İlk aşamada veri setindeki tüm öznelikler kullanılarak 4 farklı makine öğrenmesi algoritmasının sınıflandırma başarıları test edilmiştir. Sınıflandırma için kullanılan modellerin performansları temelde karmaşıklık matrisleri ve bunun üzerinden elde edilen alt metriklere göre raporlanmıştır. Tablo 3'te verilmiştir.

Tablo 3. Karmaşıklık matrisi.

Tahmin Edilen Sınıf	Gerçek Sınıf		
		OSB'li	OSB'li değil
	OSB'li	DP	YP
OSB'li değil	YN	DN	

Sınıflandırıcı tarafından OSB'li olduğu belirlenen örneğin gerçek sınıfı da OSB'li ise bu durum Doğru Pozitif (DP), gerçek sınıfı OSB'li değil ise Yanlış Pozitif (YP) olarak adlandırılır. Benzer şekilde sınıflandırıcının OSB'li olmadığını belirlediği örneğin gerçek sınıfı da OSB'li değil ise bu durum Doğru Negatif (DN), gerçek sınıfı OSB'li ise Yanlış Negatif (YN) olarak adlandırılır. Bu değerler kullanılarak hesaplanan performans metrikleri ve formülleri Tablo 4'te verilmiştir.

Tablo 4. Model performans metrikleri.

Metrik	Matematiksel İfadesi
Doğruluk	$(DP + DN) / (DP + YP + YN + DN)$
Kesinlik	$DP / (DP + YP)$
Duyarlılık	$DP / (DP + YN)$
F1 Skoru	$2 * Kesinlik * Duyarlılık / (Kesinlik + Duyarlılık)$

Öznitelik seçimi yapmadan ve 5 farklı öznitelik seçimi uygulanarak veri seti üzerinde KNN, LR, NB ve DVM algoritmaları ile sınıflandırmalar gerçekleştirilmiştir. Sınıflandırma işlemlerinde 5 kat çapraz doğrulama uygulanmış ve skorların ortalamaları, Tablo 4'te belirtilen metrikler ile raporlanmıştır. Öznitelik seçimi uygulanmadan gerçekleştirilen bu sınıflandırmalar sonucunda elde edilen bulgular Tablo 5'te verilmiştir.

Tablo 5. Öznitelik seçimi yapılmadan elde edilen sonuçlar.

Öznitelik Sayısı	Sınıflandırıcı (%)	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1 Skoru (%)
18 (tamamı)	KNN	84,3601	83,8862	87,2037	86,6666
	NB	86,7298	90,0473	86,1904	87,7594
	LR	89,5734	88,6255	86,1904	87,0959
	DVM	83,8862	86,2559	83,3333	85,6714

Veri setinde yer alan 17 girdi ve 1 hedef değişken, toplamda 18 öznitelik içermektedir. Ön işleme aşamasını takiben gerçekleştirilen öznitelik seçimi adımında, kapsamlı bir "feature selection" (öznitelik seçimi) stratejisi kullanılmıştır. Bu strateji, belirli özelliklerin çıkarılması ve ardından modellerin eğitilmesi aşamasında önem taşımaktadır. Veri setindeki önemli öznitelikleri belirlemek ve modelin genel performansını artırmak amacıyla çeşitli öznitelik seçim yöntemleri içermektedir. KNN, DVM, LR ve NB gibi çeşitli sınıflandırma modelleri, seçilen önemli özniteliklerle birlikte oluşturulmuştur. Her bir durumda seçilen öznitelikler ve bu öznitelikler üzerinde gerçekleştirilen sınıflandırmalara ait bulgular Tablo 6'da detaylı olarak sunulmuştur. Tablo 6'da ki sonuçlar incelendiğinde, seçilen 5 adet öznitelik üzerinde DVM ve LR sınıflandırıcı ile elde edilen başarı oranı %100 olarak elde edilmiştir. Farklı sınıflandırmalarda seçilen en az öznitelik sayısı 5 olup, bu sayıda öznitelikle elde edilen en yüksek doğruluk oranı KNN algoritması ile %99,52 olarak belirlenirken, NB algoritması ile %96,68 olarak elde edilmiştir. Sınıflandırmalar sonucunda, öznitelik seçimi yapılan durumlar ile yapılmayan durumlar arasındaki en yüksek doğruluk değerleri özet olarak Tablo 7'de sunulmuştur.

Tablo 6. Öznitelik seçimi ile elde edilen sonuçlar.

Öznitelik	Sınıflandırıcı	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru
Tümü	NB	86,7298	90,0473	86,1904	87,7594
	KNN	84,3601	83,8862	87,2037	86,6666
	DVM	83,8862	86,2559	83,3333	85,6714
	LR	89,5734	88,6255	86,1904	87,0959
1-4-11-12-13	NB	96,6824	93,8388	92,8909	93,4529
	KNN	99,5260	1	1	99,9052
	DVM	1	1	1	1
	LR	1	1	1	1
2-5-6-7-8-9-11	NB	95,2606	91,4691	91,9431	91,7436
	KNN	95,2606	92,4170	91,9431	92,1236
	DVM	96,6824	91,9431	91,4691	91,7436
	LR	97,1563	93,3649	91,4691	92,1236
11-12-13-14-15-16	NB	78,1990	82,4644	76,6666	71,4470
	KNN	99,0521	1	99,0476	99,24080
	DVM	1	1	1	1
	LR	1	1	1	1
2-5-6-7-9-11-13-16	NB	90,6303	91,9431	89,2606	94,1173
	KNN	92,4170	90,9952	90,5213	90,1308
	DVM	95,7345	93,3649	91,4691	91,8370
	LR	96,2085	93,3649	91,9431	92,1214

Tablo 7. Elde edilen en iyi doğruluk değerleri.

Öznitelik	NB	KNN	DVM	LR	Ortalama
Tümü	86,7298	84,3601	83,8862	89,5734	86,54022
1-4-11-12-13	96,6824	99,5260	1	1	99,0521
2-5-6-7-8-9-11	95,2606	95,2606	96,6824	97,1563	96,089975
11-12-13-14-15-16	78,1990	99,0521	1	1	94,312775
2-5-6-7-9-11-13-16	90,6303	92,4170	95,7345	96,2085	93,747575
Ortalama	89,50042	94,12316	95,26062	96,58764	

Sonuçlar detaylı bir analiz ile incelendiğinde, öznitelik seçiminin sınıflandırma algoritmalarının performansına olan etkisi açıkça belirginleşmektedir. Özellikle, öznitelik seçimi yapılan tüm durumlarda, elde edilen sınıflandırmaların başarısı, öznitelik seçimi yapılmayan durumlara kıyasla anlamlı bir şekilde daha yüksektir. Dört farklı sınıflandırıcı arasında yapılan karşılaştırmada, başarı değeri bakımından en yüksek performans DVM ve LR modeli ile elde edilmiştir ve bu durum %100 başarı oranıyla ifade edilmiştir. İkinci sırada ise KNN %99,52 başarı oranıyla yer alırken, üçüncü sırayı %96,68 başarı oranı ile NB almıştır.

3.1 Tartışma

Bu çalışmada otizm spektrum bozukluğu (OSB) tanısında farklı makine öğrenimi modellerinin ve öznitelik seçme yöntemlerinin etkinliğini karşılaştırmalı olarak inceleyerek, bu modellerin klinik uygulamalarda potansiyelini değerlendirmeyi amaçlamaktadır. Elde edilen bulgular, sınıflandırıcıların performansının, öznitelik seçiminin model doğruluğunu önemli ölçüde artırabileceğini göstermektedir. Özellikle, Destek Vektör Makineleri

(DVM), Rastgele Orman (RO) ve Yapay Sinir Ağları (YSA) gibi modellerin yüksek doğruluk oranlarına sahip olduğu gözlemlenmiştir. Literatürdeki benzer çalışmalarla kıyaslandığında, kullanılan öznelik azaltma ve sınıflandırma yöntemlerinin performansı belirgin şekilde öne çıkmaktadır. Çalışmanın en önemli katkılarından biri, veri setinde yer alan 17 girdi ve 1 hedef değişken, toplamda 18 öznelik üzerinden gerçekleştirilen analizlerin, OSB tanısında hangi özneliklerin en kritik olduğunu göstermesidir. Öznelik seçiminin, sınıflandırma başarısını artırmada kilit rol oynadığı ve minimum sayıda öznelik ile yüksek doğruluk oranlarına ulaşılabildiği görülmektedir. Bu durum, OSB tanısında daha az veri ile daha yüksek başarı elde edilebileceğini ve dolayısıyla tanı süreçlerinin optimize edilebileceğini ortaya koymaktadır. Bu bulgu, özellikle büyük veri setleriyle çalışmanın zor olduğu klinik ortamlar için önemli bir pratik katkı sunmaktadır. Çalışmanın sınırlılıkları arasında, kullanılan veri setinin büyüklüğünün sınıflandırıcıların performansını nasıl etkilediği konusunda sınırlı bilgi sunması sayılabilir. Özellikle geniş veri setlerinde bu modellerin doğruluk oranlarının ne kadar korunabileceği gelecekteki çalışmalar için bir araştırma konusu olabilir. Ayrıca, farklı demografik gruplar ve yaş aralıklarında makine öğrenmesi modellerinin etkinliğini incelemek, OSB tanısında daha kapsayıcı ve genellenebilir sonuçlar elde edilmesine katkı sağlayabilir. OSB tanısında kullanılan makine öğrenmesi modellerinin etkinliğini artırmak için öznelik seçiminin kritik bir adım olduğunu ve sınıflandırıcı modellerin doğruluğunu önemli ölçüde etkilediğini göstermektedir. Çalışmada elde edilen yüksek doğruluk oranları, bu yöntemlerin klinik uygulamalarda etkili bir şekilde kullanılabilirliğini göstermekte olup, gelecekteki araştırmalar için önemli bir temel oluşturmaktadır. Özellikle, farklı makine öğrenimi algoritmalarının OSB tanısında daha fazla araştırılması ve bu algoritmaların klinik uygulamalarda daha geniş bir şekilde test edilmesi, alanın gelişimi açısından kritik olacaktır.

Bu çalışmada elde edilen bulguların, aynı veri setini kullanılan literatürdeki güncel çalışmaların bulguları ile karşılaştırılması Tablo 8’de verilmiştir.

Tablo 8. Bulguların literatürdeki çalışmalar ile karşılaştırılması.

Referans	Yöntem	En İyi Doğruluk (%)
1 de Campos Souza & Guimaraes (2018)	YSA	100
2 Akyol ve Karaci (2018)	LR	97,33
3 Shuvo vd (2019)	RO	96
4 Misman vd (2019)	DNN, DVM	95,24
5 Raj & Masood, (2020)	NB, DVM, LR, KNN, CNN, YSA	99,53
6 Metlek ve Kayaalp (2020)	YSA, DK SBS	100
7 Thabtah & Peebles (2020)	NB, LR	97,94
8 Mokni ve Haoues (2020)	Decision Tree, DVM	99,50
9 Khatun vd (2021)	DVM	85
10 Santos (2021)	KNN, DVM, RO	79,35
11 Saihi & Alshraideh (2021)	RO, C4.5 Tree-, YSA	99
12 Saran & Pirouz (2021)	KNN, RO, DVM, LR	73
13 Alaika & Alamsyah (2022)	GSA, DVM, CFS	96,61
14 Dedgaonkar vd (2022)	KNN, CNN	100
15 Uddin (2023)	RO, MNB, BNB, DVM, NB	100
16 Praveena vd (2023)	Adaboost given, RO	100
17 Pottem vd (2023)	NB, KNN	97,47
18 Bidwe vd (2023)	CNN	87,99
19 Bu çalışma	DVM, LR	100

Tablo 8’de sunulan verilere göre, son altı yılda gerçekleştirilen benzer çalışmalarda en yüksek doğruluk %100 iken en düşük doğruluk %73 olarak kaydedilmiştir. Bu çalışmada elde edilen en yüksek doğruluk değeri de %100’e ulaşmıştır. Bu çalışmada, öznelik seçimi ve sınıflandırıcıların kullanımıyla elde edilen sonuçların, benzer çalışmalara kıyasla yüksek bir performans sergilediği tespit edilmiştir. Tabloda yer alan diğer çalışmalar veri setindeki özneliklerin tamamı üzerinde işlem yaparken, gerçekleştirilen çalışmada önce öznelik seçimi gerçekleştirilmiş, daha az öznelik kullanarak daha performanslı sonuçlar elde edilmeye çalışılmıştır. Özellikle,

Destek Vektör Makinesi ve Lojistik Regresyon Modellerinin 5 öznitelik ile %100 doğruluk sağladığı diğer benzer çalışmalardan daha başarılı sonuçlar verdiği gözlemlenmiştir. Bu sonuç, özellikle OSB veri setindeki öznitelik seçimi yoluyla daha etkin sonuç elde etme konusuna odaklı araştırmalar için önemli bir bulgudur. Literatürdeki benzer çalışmaların bulguları incelendiğinde bu çalışmada kullanılan yöntemlerin diğer çalışmalarda rekabet edebilecek düzeyde ve genelinden daha yüksek skorlar sağladığı görülmektedir. Benzer çalışmalarda aynı yüksek doğruluk seviyelerine ulaşmasına rağmen, bu çalışmada daha az sayıda öznitelik kullanılarak %100'e ulaşan başarı sağlanmıştır. Bu durum, daha az ve etkin özniteliklerle yüksek başarıya ulaşabildiğini göstermektedir. Bu çalışmanın özellikle öznitelik seçimi yöntemlerinin ve farklı sınıflandırıcıların kombinasyonunun ne kadar etkili olabileceğini göstermektedir. 5 öznitelikle %100 doğruluk elde edilmesi, öznitelik seçiminin önemini ve doğru yöntemlerin kullanılmasının sonuçlara olan etkisini açıkça ortaya koymaktadır.

4. Sonuç

Erken teşhis ve etkili müdahaleler, OSB ile ilişkili zorlukların üstesinden gelmede hayati önem taşımaktadır. Erken tanı, müdahalelerin ve terapilerin etkinliğini önemli ölçüde artırabileceğinden dolayı son derece kritik bir konudur. Erken teşhis, müdahalelerin ve terapilerin etkisini belirgin şekilde artırma potansiyeline sahiptir ve bu nedenle büyük önem taşımaktadır. OSB veri seti üzerinde gerçekleştirilen öznitelik seçimi ve sınıflandırıcıların performansı veri setindeki 17 girdi ve 1 hedef değişken, toplamda 18 öznitelik üzerinde yapılan analizler, öznitelik seçiminin sınıflandırma modellerinin başarısına olan etkisini açıkça ortaya koymuştur. Öznitelik seçimi yapılmadan önce tüm özniteliklerin kullanılmasıyla elde edilen sonuçlar, farklı sınıflandırma modelleri için kabul edilebilir doğruluk oranlarına işaret etmiştir. Ancak, öznitelik seçimi yapıldığında, sınıflandırma modellerinin başarı oranlarında belirgin bir artış gözlemlenmiştir. Özellikle, DVM ve LR modelleri, minimum sayıda öznitelikle %100 doğruluk oranına ulaşarak diğer modellerden ayrılmıştır. Literatürdeki benzer çalışmaların doğruluk değerleriyle karşılaştırıldığında, bu çalışmanın kullanılan yöntemler ve modeller açısından daha yüksek başarı elde ettiği görülmüştür. Öznitelik seçme yöntemlerinin etkin kullanımıyla daha az sayıda öznitelikle yüksek doğruluk oranlarına ulaşılması, özellikle veri setindeki önemli özniteliklerin belirlenmesi ve model performansının artırılmasında önemli bir artış görülmüştür. Bu çalışma öznitelik seçimi yöntemlerinin ve farklı sınıflandırıcıların kombinasyonunun sınıflandırma modellerinin performansını nasıl etkileyebileceği gösterilmiştir. Özellikle, doğru öznitelik seçimi ve uygun sınıflandırıcıların seçimiyle elde edilen %100 doğruluk oranı, benzer çalışmalardan daha yüksek bir başarı düzeyini temsil etmektedir. Bu nedenle, öznitelik seçimi ve sınıflandırma modellerinin dikkatli bir şekilde seçilmesi ve uygulanması özellikle karmaşık veri setleri üzerinde yüksek doğruluk ve performans sağlayabileceğini göstermektedir. Bu bulgular, veri madenciliği ve sınıflandırma alanında çalışan araştırmacılar için değerli bir rehberlik sunmaktadır. Özellikle karmaşık veri setleri üzerinde yüksek doğruluk ve performans sağlamak için öznitelik seçimi ve sınıflandırma modellerinin dikkatlice seçilmesi ve uygulanması gerektiği konusunda araştırmacılara yönlendirme sağlanabilir. Bu şekilde, gelecekteki çalışmaların daha etkili ve güvenilir sonuçlar elde etmesinde katkı sağlayacak niteliktedir.

Çalışmanın bazı sınırlamaları da göz önünde bulundurulmalıdır. İlk olarak, çalışma yalnızca belirli bir veri seti üzerinde gerçekleştirilmiştir; dolayısıyla, elde edilen bulgular farklı veri setlerine genellenebilir mi sorusu araştırılması gereken bir konudur. Farklı veri setleri üzerinde benzer sonuçların elde edilip edilemeyeceği, gelecekteki çalışmaların ana odaklarından biri olmalıdır. Ayrıca, öznitelik seçimi yöntemlerinin etkinliği, veri setinin özelliklerine ve yapısına bağlı olarak değişebilir. Bu nedenle, çalışmada kullanılan öznitelik seçimi yöntemlerinin, farklı veri türleri ve boyutları üzerinde nasıl performans göstereceği daha fazla araştırılmalıdır. Bir diğer sınırlama ise, çalışmada yalnızca belirli sayıda sınıflandırıcı modelin değerlendirilmiş olmasıdır. İleride yapılacak çalışmalar, farklı sınıflandırıcı modelleri veya derin öğrenme algoritmalarını da inceleyerek, bu modellerin OSB veri setleri üzerindeki performanslarını değerlendirebilir. Son olarak, %100 doğruluk oranına ulaşan modellerin genelleme yetenekleri, özellikle daha büyük ve daha karmaşık veri setleri üzerinde test edilmelidir. Modelin aşırı öğrenme (overfitting) yapma riski dikkate alınarak, genelleme yeteneği daha geniş çaplı analizlerle değerlendirilmelidir.

Bu çalışmanın bulgularını daha da genişletmek için çeşitli yaklaşımlar önerilebilir. İlk olarak, daha geniş ve çeşitlendirilmiş veri setleri üzerinde ek analizler yapılarak, farklı öznitelik seçimi yöntemlerinin ve sınıflandırıcıların performansı test edilebilir. Böylelikle, bulguların genellenebilirliği artırılabilir. Ayrıca, derin öğrenme teknikleri kullanılarak daha karmaşık modellerin incelenmesi, özellikle büyük veri setlerinde doğruluk oranlarının daha da iyileştirilmesine katkı sağlayabilir. İleride yapılacak çalışmalar, özellikle bu tür modellerin hesaplama maliyetlerini ve uygulanabilirliğini de göz önünde bulundurmalıdır. Çalışmanın bulguları, özellikle bazı öznitelik kombinasyonlarıyla önemli başarılar elde edildiğini göstermektedir. Tablo 8'de görüldüğü gibi, öznitelik seçimi yapıldıktan sonra DVM ve LR modelleri %100 doğruluk oranına ulaşmış ve diğer modellerden belirgin şekilde ayrılmıştır. Özellikle, "1-4-11-12-13" ve "11-12-13-14-15-16" öznitelik kombinasyonları kullanıldığında, DVM ve LR modellerinin doğruluk oranları en yüksek seviyeye ulaşmıştır. Ortalama doğruluk

oranları incelendiğinde, LR modeli %96,59, DVM modeli ise %95,26 doğruluk ile öne çıkmaktadır. Bu sonuçlar, öznelik seçiminin ve sınıflandırıcıların dikkatli bir şekilde belirlenmesinin sınıflandırma performansını nasıl artırabileceğine dair önemli bir kanıt sunmaktadır.

Kaynaklar

- [1] Korkmaz B. Otizm: Klinik ve nörobiyolojik özellikleri erken tanı tedavi ve bazı güncel gelişmeler çağrılı Yazar. Türk Pediatri Arşivi 2010; 45(12): 1-8.
- [2] Çattik EO, Yetkin Aİ, Diken İH. Erken çocukluk döneminde otizm spektrum bozukluğunda aile merkezli müdahaleler. Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi 2020; 21(3): 589-610.
- [3] Volkmar FR, McPartland JC. From Kanner to DSM-5: Autism as an evolving diagnostic concept Annu Rev Clin Psychol, 2014 ;10: 193-212.
- [4] Akkuş PZ, Saygan BB, Bahadır Eİ, Cak T, Özmert E. Otizm spektrum bozukluğu tanısı ile yaşamak: Ailelerin deneyimleri. Turkish Journal of Pediatric Disease 2021; 15(4): 272-279.
- [5] Atalay M, Çelik E. Büyük veri analizinde yapay zekâ ve makine öğrenmesi uygulamaları- Artificial intelligence and machine learning applications in big data analysis. Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi 2017; 9(22): 155-172.
- [6] Tamur Y, Celasin NŞ. Aile sağlığı merkezlerinde görev yapan sağlık profesyonellerinin otizm konusundaki bilgi ve erken tanıya yönelik görüşleri. Turkish Journal of Family Medicine and Primary Care 2022; 16(3): 517-531.
- [7] Başkak ET, Aral N. Türkiye’de Otizm tanısı almış bireylerin aileleri ile yapılmış tezlerin ve makalelerin incelenmesi (2000-2021). İzmir Sosyal Bilimler Dergisi 2021; 3(2): 213-224.
- [8] Çelik S, Çeşmeli MŞ. Çocuklar için Otizm Spektrum bozukluğunun veri madenciliği yöntemleri ile analizi. Acta Infologica 2021; 5(1): 167-186.
- [9] Demirhan,A. Performance of machine learning methods in determining the autism spectrum disorder cases. Mugla. Journal Of Science And Technology 2018; 4(1): 79-84.
- [10] Metlek S, Kayaalp K. Otistik Spektrum bozukluğunun makine öğrenme algoritmaları ile tespiti. Int J Intell. Syst International.: Theory and Applications 2020; 3(2): 60-68.
- [11] Thabtah F, Peebles D. A new machine learning model based on induction of rules for autism detection. Health Inf J 2020; 26(1): 264-286.
- [12] Akyol K, Karaci A. A study on autistic spectrum disorder for children based on feature selection and fuzzy rule. International congress on engineering and life science 2018: 804-807.
- [13] De Campos Souza PV, Guimaraes AJ. Using fuzzy neural networks for improving the prediction of children with autism through mobile devices. Symposium on computers and communications 2018; 01086-01089.
- [14] Saihi A, Alshraideh H. Development of an autism screening classification model for toddlers. Conference on Machine Learning & Applications 2021; 2110.01410.
- [15] Shuvo SB, Ghosh J, Oyshi AS. A data mining based approach to predict autism spectrum disorder considering behavioral attributes ICCCNT 2019; 1-5.
- [16] Bidwe RV, Mishra S, Bajaj S. Performance evaluation of Transfer Learning models for ASD prediction using non-clinical analysis. Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing 2023; 474-483.
- [17] Raj S, Masood S. Analysis and detection of autism spectrum disorder using machine learning techniques. Procedia Comput Sci 2020; 167: 994-1004.
- [18] Alaiika L, Alamsyah A. Optimization of accuracy to autism spectrum disorder identification for children using support vector machine and correlation-based feature selection. Journal of Advances in Information Systems and Technology 2022; 4(1): 1-12.
- [19] Saran PK, Pirouz M. EEG analysis for predicting early autism spectrum disorder traits. Proceedings of the Future Technologies Conference 2021; 2: 658-675.
- [20] Mismar MF, Samah AA, Ezudin FA, Majid HA, Shah ZA, Hashim H, Harun MF. Classification of adults with autism spectrum disorder using deep neural network. In2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS); 2019; Sep 19; Malaysia: 29-34.
- [21] Khatun MA, Ali MA, Ahmed MR, Noori SRH, Sahayadhas A. Empirical study of computational intelligence approaches for the early detection of autism spectrum disorder. Intelligent Computing and Applications: Proceedings of ICICA 2019; 2021; 161-170.
- [22] Santos DF. Analyzing the predictive power of machine learning models for autism detection. SciELO Preprints 2023.
- [23] Mokni R, Haoues M. Deep Classifier Model for Autism Spectrum Disorder Prediction. Knowledge Innovation Through Intelligent Software Methodologies, Tools and Techniques. IOS Press; 2020: 138-49.
- [24] Dedgaonkar S, Sachdeo R, Godbole S. Hybrid feature vector for screening of autistic people using deep learning. International Journal of Intelligent Engineering & Systems 2022; 15(1).
- [25] Uddin KMM. A machine learning approach to predict autism spectrum disorder (ASD) for both children and adults using feature optimization. Network Biology 2023; 13(2): 37.

- [26] Praveena KN, Mahalakshmi R, Manjunath C, Zubair AF, Karthikeyan P. Optimized Feature selection and transformations for early stage prediction of autism using supervised machine learning models. *MECS* 2023; 6: 73-89.
- [27] Pottam A, Sirisha G, Chaitanya RK. Prediction of autism spectrum disorder from high dimensional data using machine learning techniques. 2023;28:261.
- [28] Günay M, Alkan A. Classification of EMG signals by spectral methods and SVM classifier. *Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi* 2016; 13(2): 63-80.
- [29] Balcı M, Alkan A. Identification of wart treatment evaluation by using optimum ensemble based classification techniques. *Biomed Signal Process Control* 2024; 95: 106491.
- [30] Thabtah F, Peebles D. A new machine learning model based on induction of rules for autism detection. *Health Inf J* 2020; 26(1): 264-286.
- [31] Autism screening data for toddlers 2018. <https://www.kaggle.com/datasets/fabdelja/autism-screening-for-toddlers>.
- [32] Çelik Ö. A Research on Machine Learning Methods and Its Applications. *J Educ Technol Online Learn*. 2018;1(3):25-40.
- [33] Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. *Proceedings Of The Fifth Annual Workshop On Computational Learning Theory* 1992; 144-152.
- [34] Bone D, Bishop SL, Black MP, Goodwin MS, Lord C, Narayanan SS. Use of machine learning to improve autism screening and diagnostic instruments: Effectiveness, efficiency, and multi-instrument fusion. *J Child Psychol Psychiat.*; 57(8): 927-937.
- [35] Pisner DA Schnyer DM. Chapter 6-Support vector machine, *Machine Learning*. A. Mechelli & S. Vieira (Ed.). Academic Press, 2020; 101-121.
- [36] Nick TG, Campbell KM. Logistic regression, *Topics in Biostatistics*. W. T. Ambrosius (Ed.). Humana Press, 2007; 273-301.
- [37] Huang X, Jin G, Ruan W. Naive bayes, *Machine Learning Safety*. X. Huang, G. Jin, & W. Ruan (Ed.). Springer Nature 2023; 95-102.
- [38] Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: A new perspective. *Neurocomputing* 2018; 300: 70-79.
- [39] Fernandez A, Garcia S, Herrera F, Chawla NV. SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *J Artif Intell Res*, 2018; 61: 863-905.
- [40] Elreedy D, Atiya AF, Kamalov F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Mach learn* 2024; 113(7): 4903-4923.