

*Araştırma makalesi*

## Naive Bayes ile Otistik Spektrum Bozukluğu Tanısı Koyma

Mustafa SU <sup>1\*</sup>, Hidayet TAKCI <sup>1</sup>

<sup>1</sup> Cumhuriyet Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, htakci@cumhuriyet.edu.tr

\* Sorumlu Yazar; su.mustafa@hotmail.com

Gönderme tarihi: 25/12/2022

Kabul tarihi: 05/05/2023

### ÖZET

Otizm kişinin çevresiyle uygun ilişkiler kuramadığı gelişimsel bir bozukluktur. Otizmin erken teşhisi kişinin erken yaşta eğitilip çevresiyle olan ilişkisinin düzenlenmesinde hayati önem taşımaktadır. Bu çalışmada, otizm teşhisinin geç konulması sebebiyle otizm teşhisi konulan kişilerin yaşadığı sorunların çözümüne katkı sağlayabilmek için Naive Bayes Makine Öğrenmesi algoritmasıyla hastalığa erken tanı konulması amaçlanmıştır. Çalışma analizinin gerçekleştirilebilmesi için veri setinde otizmlili insanların bebeklik, ergenlik ve yetişkinlik dönemleri ele alınmıştır. Veriler üzerinde; aykırılıkların ayıklanması, eksik verilerin doldurulması, öznitelik ağırlığına göre öznitelik seçimi yapılması ve parametre optimizasyonu sonrası eğitilen modellerin başarı oranları ile ham veri üzerinden parametre optimizasyonu yapılmadan elde edilen modellerin başarı oranları karşılaştırılmıştır. Yapılan çalışmalar neticesinde parametre optimizasyonu yapılmamış veri setine oranla; yetişkin veri setinde modelin başarısı % 2.27, ergen veri setinde %7.01 ve çocuk veri setinde %3.78 performans artışı gözlemlenmiştir. Bu çalışma, veri ön işleme teknikleri ve parametre optimizasyon işlemleri sonrası başarı oranlarının arttığını göstermektedir.

**Anahtar Kelimeler:** Otizm; Naive Bayes; Veri Ön İşleme; Parametre Optimizasyonu

## Diagnosing Autistic Spectrum Disorder with Naive Bayes

### ABSTRACT

Autism is a developmental disorder in which a person cannot establish appropriate relationships with their environment. Early diagnosis of autism is of vital importance in educating the person at an early age and regulating his relationship with his environment. In this study, it is aimed to diagnose the disease early with the Naive Bayes Machine Learning algorithm in order to contribute to the solution of the problems experienced by people diagnosed with autism due to the late diagnosis of autism. In order to carry out the study analysis, the infancy, adolescence and adulthood periods of people with autism were discussed in the data set. These periods were categorically examined separately in the data set. The success rates of the models trained after removing outlier data, filling in

missing data, selecting features according to the feature weight and parameter optimization, and the success rates of the models obtained from raw data and without parameter optimization were evaluated. In the study with the adult data set, 2.27% performance increase was observed in the success of the model, 7.01% in the study with the adolescent data set, and 3.78% in the study conducted with the child data set. This study shows that success rates are increased after data preprocessing techniques and parameter optimization processes.

**Keywords:** Autism; Naive Bayes; Data Preprocessing; Parameter Optimization

## 1. GİRİŞ

Otizm genellikle yaşamın ilk yıllarında başlayan ve ölene kadar devam eden, bireyin çevresiyle hem sözel hem de sözel olmayan şekilde uygun ilişki kuramaması şeklinde bireyde iletişim ve etkileşim konularında kısıtlamalara sebep olan gelişimsel bir bozukluktur. Otizmli bireyler yaşamın farklı alanlarında sosyalleşme, iletişim ve davranış konularında bazı sorunlar yaşarlar. Bu sorunlar sebebiyle genellikle sözlü iletişim ve arkadaşlık kurma konusunda sıkıntıları bulunur. Otizm kaynaklı gelişim bozukluğu bireyde duygusal durum değişikliklerine sebep olabilir ve kendilerini etkili ifade edemeyebilirler. Kendilerini ifade edememelerinden dolayı öfke nöbetlerine yakalanır ve nasıl sakinleşecekleri konusunda kendilerini kontrol edemezler. Devrik ve karmaşık cümleler ile günlük hayatta kullanılan mecazi ifadeleri anlayamazlar. Ancak sosyal etkileşim, bilinçli ebeveynler ve eğitim desteği ile bu sorunlarla başa çıkmayı öğrenebilirler (Özkan ve ark., 2015). Hastaların eğitimine erken başlanması hastalığın tedavisi ve bireylerin gelişimi açısından oldukça önemlidir. Ancak otizmin teşhisi zor ve uzun bir süreç olduğundan dolayı tedavi sürecinin başlanmasına geç kalınabilmektedir. Hastalığın teşhisi sürecindeki gecikmeleri ve belirsizlikleri engellemek için son dönemde sıklıkla makine öğrenmesi teknikleri kullanılmaktadır.

Bu çalışma kapsamında da üç ayrı yaş grubuna ait veri seti üzerinde Naive Bayes algoritmasının orijinal verideki başarı oranları ile veri ön işleme ve parametre optimizasyonu sonrasındaki başarı oranları karşılaştırılmış ve raporlanmıştır. Bu yöntemle eğitilen modellerin başarısı yüzde 2 ile yüzde 8 arasında artış göstermiştir. Çalışma, parametre optimizasyonu ve veri ön işleme tekniklerinin otizm tespit oranlarını artırıp artırmadığını görmek üzere yapılmıştır.

Literatürde farklı veri madenciliği algoritmaları ile Otizm teşhisi konusunda yapılan araştırmalar ve algoritmaların bazılarını şu şekilde özetlemek mümkündür:

2020 yılında Sedat Metlek tarafından hazırlanan “Otistik Spektrum Bozukluğunun Makine Öğrenme Algoritmaları ile Tespiti” adlı çalışmada bahsedildiğine göre nöro-gelişimsel bir gelişme bozukluğu olan Otizmin erken teşhisi için 12-36 ay arasındaki çocuklara yüksek

doğrulukta OSB teşhisi için yardımcı olabilecek bir karar destek yazılımı geliştirilmiştir. Yapılan çalışmada altı farklı makine öğrenmesi algoritması ile otizm teşhisi yapılmıştır. Çalışmaya göre; gözetimli ve gözetimsiz modellerin test edilmesi sonucunda gözetimli öğrenme algoritmalarının daha başarılı olduğu gözlemlenmiştir (Metlek ve Kayaalp, 2020).

Otizm teşhisi konusunda başarı sağlayan bir diğer çalışma 2019 yılında Cho ve arkadaşları tarafından yayınlanan “Automatic Detection of Autism Spectrum Disorder in Children Using Acoustic and Text Features from Brief Natural Conversations.” başlıklı çalışmadır. Bu projenin amacı, sonunda doktorlar ve okullar gibi gerçek dünya ortamlarında OSB'nin ön taraması için kullanılacak çok kısa, genel ve doğal konuşmalara dayanan otomatik bir OSB saptama sistemi geliştirmektir. Bu kapsamda Gradient tabanlı bir algoritma ile çalışma gerçekleştirilmiş ve teşhis konusunda %76 oranında başarı elde edilmiştir (Cho ve ark., 2019).

Jaber Alwidian, Ammar Elhassan ve Ghnemat'ın çalışmasında ise bireyin otizmi olup olmadığı ilişkilendirme sınıflandırması tekniği kullanılarak yedi farklı algoritma ile test edilmiştir. Aralarındaki korelasyonlar incelenmiş ve en yüksek doğruluk oranı ile çalışan algoritma tespit edilmiştir. Bu çalışmanın sonucu olarak Birliktelik Kuralları Algoritmasına Dayalı Ağırlıklı Sınıflandırma (WCBA) algoritması en yüksek doğruluğu vermiştir (Alwidian ve ark., 2020).

Suman Raj ve Sarfaraz Masood tarafından 2019 yılında Uluslararası Sayısal Zekâ ve Veri Bilimi Konferansı'nda yayınlanan “Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques” çalışması ise otizm gelişim bozukluğunun üç ayrı gelişim evresine göre analizini ele almaktadır. Bu çalışmada Naive Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network ve Convolutional Neural Network uygulamalarının üç ayrı yetişkinlik düzeyinde performansı ölçülmüştür. Bu çalışma sonucunda Convolutional Neural Network uygulamasının doğruluk oranı daha yüksek çıkmıştır ve teşhis için en uygun tekniğin Convolutional Neural Network olduğu gözlemlenmiştir (Raj ve Masood, 2020).

M. S. Mythili ve A. R. Mohamed Shanavas “Sınıflandırma Teknikleri Kullanılarak Otizm Spektrum Bozuklukları Üzerine Bir Çalışma” isimli bir makale yayınlamışlardır. 2014 yılında yayınlanan bu makalede otizm düzeyleri, veri madenciliği ve sınıflandırma algoritmalarıyla tespit edilmeye çalışılmıştır. Çalışmanın amacı otizm problemini ve derecesini tespit etmektir. Böylece çocukların okul performansının artırılması hedeflenmiştir. Otistik spektrum

bozukluğunun (OSB) sınıflandırılması için bulanık bilişsel harita optimizasyonunun verimliliğinden bahsedilmiştir (Mythili ve Shanavas, 2014).

Hailong Li ve arkadaşları tarafından yayınlanan “Optimize Edilmiş Makine Öğrenimi Modelleri ve Kişisel Karakteristik Verilerle Otizm Teşhisini Geliştirmek” adlı çalışma günümüz klinik yöntemlerin haricinde kişisel verilere de odaklanmıştır. Kamuya açık bir veri tabanı üzerinde yaptıkları çalışmada bireyin yaş, cinsiyet, IQ seviyesi gibi kişisel verileri de kullanılmış ve denetimli makine öğrenimi deneyleri yapılmıştır. Dokuz ayrı denetimli makine öğrenimi modelinin test edildiği bu çalışmada bireyin karakteristik özelliklerinin sınıflandırma performansı gözlemlenmiştir. En iyi performansı öncelikle yapay sinir ağı modeli ve sonrasında k-en yakın komşu algoritması vermiştir (Parikh ve ark., 2019).

Devika Varshini G ve Chinnaiyan R, 2020 yılında “Otizm Spektrum Bozukluğunun Tahmini için Optimize Edilmiş Makine Öğrenimi Sınıflandırma Yaklaşımları” adlı çalışmalarını yayınlamışlardır. Bu çalışma bebeklik ve yetişkinlik dönemindeki bireylerin gösterdikleri otizm özelliklerinin sınıflandırma performansını ölçmek amacıyla yapılmıştır. Psikologlar ve davranış bilimcilerin oluşturduğu puanlama yönteminden yararlanılarak bu proje geliştirilmiştir. Optimize edilmiş makine öğrenimi metotlarının analiz edildiği bu çalışmada Lojistik Regresyon ve Random Forest gibi algoritmalar ile otizm bozukluğunun tahmininin eleştirisi ve analizi yapılmıştır. Bu çalışmada KNN algoritması en yüksek doğruluk oranını vermiştir (Devika ve ark., 2020).

Fatiha Nur Büyükoflaz ve Ali Öztürk’ün “Makine Öğrenmesi Algoritmaları ile Çocuklarda Erken Otizm Teşhisi” çalışmasında; Naive Bayes, K-En Yakın Komşu ve Random Forest gibi sınıflandırma modelleri karşılaştırılmış ve bunun sonucunda en yüksek performansı Random Forest algoritması vermiştir. Naive Bayes uygulaması %96,55 başarı oranı vermiştir. Bu çalışmada düşük maliyetli, hızlı ve uygulaması kolay bir otizm tarama aracı oluşturmak için OSB riski taşıyan binlerce çocuktan elde edilen klinik verilere Makine Öğrenimini (ML) uygulamaya çalışılmıştır (Büyükoflaz ve Öztürk, 2018).

Azian Azamimi Abdullah, Saroja Rijal ve Satya Ranjan Dash’ın yayınladıkları “Otizm Spektrum Bozukluğunun Sınıflandırılmasına Yönelik Makine Öğrenimi Algoritmalarının Değerlendirilmesi” isimli çalışmalarında otizm bozukluğunun sınıflandırılması ve daha yüksek başarı oranları elde etmek için otizm sorularına yönelmişlerdir. Bu çalışmada diğer çalışmalardan farklı olarak algoritmalar ve onların doğruluk oranlarına odaklanılmamış, daha çok kullanılan veri setlerine odaklanılmış ve bu veri setleri çeşitlendirilmiştir. Denetimli

makine öğrenimi algoritmalarının ve çeşitli test yöntemlerinin denendiği bu çalışmada Lojistik Regresyon algoritmasının en yüksek doğruluğu verdiği açıklanmıştır (Abdullah ve ark., 2019).

Ayşe Demirhan'ın 2018 yılında yayınlamış olduğu olduğu "Otizm Spektrum Bozukluk Vakalarını Belirlemede Makine Öğrenme Yöntemlerinin Performansı" adlı çalışmasında Destek Vektör Makineleri, K-En Yakın Komşu ve Random Forest algoritmaları kullanılmıştır. Algoritmaların kendi içerisinde performanslarının değerlendirildiği bu çalışmada otizm bozukluğunun en hızlı ve doğru teşhisi elde edilmeye çalışılmıştır. Random Forest algoritması ile en yüksek performans elde edilmiştir. Böylece otizm bozukluğu vakalarının Random Forest algoritması ile tam başarı ile tespit edileceği analiz çıktıları arasında yerini almıştır (Demirhan, 2018).

Geçmiş çalışmalardan görüleceği üzere kimi çalışmalarda tek bir makine öğrenme algoritması kiminde birden çok makine öğrenmesi algoritması kullanılmıştır. Otizm teşhisi için kullanılan makine öğrenmesi algoritmaları denetimli öğrenme algoritmalarıdır. Çalışmalar farklı veri setleri kullansalar da hepsinin ortak amacı teşhis işlemi ve teşhisin iyileştirilmesidir. Çalışmamız da bu kapsamda ele alınmış ve bir makine öğrenmesi algoritması üzerinde parametre iyileştirme çalışmaları ile sonuç iyileştirme yapılmıştır.

## 2. MATERYALLER VE METOTLAR

### 2.1 Veri Setleri

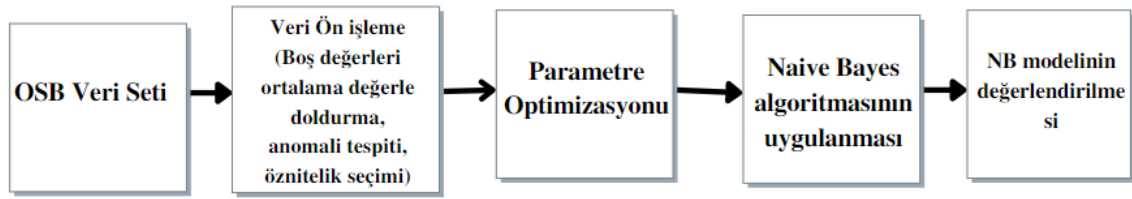
Bu çalışmada, UCI Makine Öğrenmesi Deposundan alınan üç ayrı (yetişkinlik, ergenlik ve bebeklik) veri seti kullanılmıştır. Bu veri setleri, OSB'nin teşhisinde etkili olan parametrelere sahip verileri içerir. Bu veri setinde OSB vakalarını teşhis etmede etkili olduğu kanıtlanmış 10 davranışsal özellik ve 10 bireysel özellik bulunmaktadır. Çocuk veri seti 292, ergen veri seti 104 ve yetişkin veri seti ise 704 örnek içermektedir (Anonim, 2023a; 2023b; 2023c). Bu çalışmada kullanılan veri setlerinin özellikleri ve türleri Tablo 1'de verilmiştir.

**Tablo 1.** Veri Setlerinin Özellikleri

SÜTUN	AÇIKLAMASI
1	Hasta yaşı (Yıl olarak)
2	Hasta cinsiyeti
3	Etnik köken
4	Kişinin sarılıkla doğup doğmadığı durumu
5	Herhangi bir yakın aile üyesinde osb var mı?
6	Ebeveyn, kendi, bakıcı, sağlık personeli, klinisyen vb.

SÜTUN	AÇIKLAMASI
7	Yaşanılan ülke
8	Kullanıcının bir tarama uygulaması kullanıp kullanmadığı
9	Yaş kategorisine göre seçilen tarama yöntemlerinin türü (0=bebek, 1=çocuk, 2=ergen, 3=yetişkin)
10-19	Davranışsal özelliklerle ilgili 10 soru
20	Kullanılan tarama yönteminde elde edilen nihai puan

## 2.2 Çalışma Metodolojisi



Şekil 1. Çalışma iş akış diyagramı

Şekil 1. Bu çalışmada kullanılan iş akışını göstermektedir. İlk olarak veri ön işleme teknikleri kullanılarak üç veri seti de maksimum verim alınacak hale getirilmiştir. Veri ön işleme aşamasında, aykırı veriler veri setlerinden temizlenmiş ve boş veriler o sütundaki verilerin ortalama değeri alınarak doldurulmuştur. Bu işlem her veri seti için bir kez gerçekleştirilmiştir. İkinci adım ise Parametre optimizasyonu adıdır. Bu adımda model eğitilirken ve test edilirken kullanılacak laplace yumuşatması, tek çıkışlı çapraz doğrulama, çapraz doğrulama katlama sayısı ve ağırlık seçimi gibi parametreler belirlenmiştir. Optimizasyon parametrelerinin belirlenmesinden sonra NB algoritmasının uygulanmasına geçilmiştir. Bu adım, parametre optimizasyonu adımı belirlenen parametrelerin farklı farklı kombinasyonları kullanılarak birçok kez tekrarlanmıştır.

Tablo 2. Rapidminer parametre optimizasyonu iterasyon değerleri

İterasyon	Naive Bayes Laplace Yumuşatması	Ağırlık Seçimi	Tek Çıkışlı Çapraz Doğrulama	Çapraz Doğrulama Katlama Sayısı	Doğruluk
122	Hayır	0.050	Hayır	14	0.936
123	Evet	0.060	Hayır	14	0.936
243	Evet	0	Hayır	31	0.951
124	Hayır	0.060	Hayır	14	0.936
424	Hayır	0.020	Hayır	54	0.948
125	Evet	0.070	Hayır	14	0.940
126	Hayır	0.070	Hayır	14	0.933
127	Evet	0.080	Hayır	14	0.936
128	Hayır	0.080	Hayır	14	0.933
425	Evet	0.030	Hayır	54	0.946
244	Hayır	0	Hayır	31	0.936

Tablo 2' de görüldüğü üzere modeller birçok parametre ile eğitilerek test edilmiştir. Her bir iterasyon sonucunda modelin değerlendirilmesi yapılmıştır. Bu yöntemle en iyi doğruluk oranını sağlayan parametreler belirlenerek maksimum doğruluk oranına sahip modeller elde edilebilecektir. Bu değerler rapidminer iterasyonlar ekranından alınmıştır.

### **2.2.1 Veri Ön İşleme**

Veri setindeki eksik ve tutarsız verilerin düzeltilmesi işlemi veri ön işleme aşamasında gerçekleştirilmektedir. Veri ön işleme teknikleri ile veri seti düzeltilerek makine öğrenmesi algoritmalarının performansında genellikle artış sağlanmaktadır (Kotsiantis ve ark., 2006). Aykırı verilerin belirlenmesi ve kaldırılması, eksik verilerin doldurulması, tekrarlanan verilerin kaldırılması gibi teknikler veri ön işlemede kullanılmaktadır. Bu çalışmada öklid mesafesi kullanılarak veri seti içerisindeki aykırı veriler tespit edilip filtrelenmiştir. Eksik değerler ise sütunda bulunan diğer değerlerin ortalaması alınarak doldurulmuştur.

### **2.2.2 Parametre Optimizasyonu**

Çalışmanın bu kısmında model eğitilirken farklı parametre kombinasyonları kullanılarak modeller eğitilip test edilmiştir. Aşağıda listelenen parametrelerin kombinasyonları farklı iterasyonlarda kullanılmıştır.

- Laplace yumuşatmasının kullanılma ve kullanılmama durumu
- Çapraz-doğrulama için minimum 2 maksimum 60 katlama değerlerinin kullanılması
- Öznitelik ağırlıklarının belirlenmesinde farklı yöntemler (Gini indeksi (Manek ve ark. 2017), korelasyon Jiang ve ark., 2018).)
- Öznitelik eşik değeri için minimum 0, maksimum en yüksek ağırlık değeri olan 0.155 eşik değeri

Bu parametrelerin farklı kombinasyonları ile modeller eğitilip başarı durumları gözlemlenmiş ve modelin eğitilmesinde en iyi sonucu veren parametreler belirlenmiştir. Bu kısım rapidminer üzerinde parametre optimizasyonu operatöründe paralel yürütülen işlemler ile gerçekleştirilmiştir.

### **2.2.3 Naive Bayes (NB)**

Naive Bayes denetimli öğrenme altında bir sınıflandırma algoritmasıdır. 1812 yılında Thomas Bayes tarafından bulunan koşullu olasılık teorisine dayalı olarak geliştirilmiştir.

Koşullu olasılık ile önceki olayların bilgileri göz önüne alınarak, başka olayların gerçekleşebilme olasılığını bulabiliriz (Atan, 2020). Naive bayes modeli oluşturmak için kategorileri belirli yani etiketli veriler, eğitim verisi olarak kullanılır. Bu veriler kullanılarak olasılık işlemleri yapılır ve model oluşturulur. Etiketsiz veriler bu modele verildiğinde eğitimde elde edilen olasılık değerleri kullanılarak etiketsiz verilerin etiketlerinin belirlenmesi sağlanır. Öğretilmiş veri sayısının fazla olması modelin başarısını olumlu olarak etkilemektedir.

### 2.2.4 Eğitilen Modellerin Değerlendirilmesi

Eğitilen modellerin performans değerlendirilmesinde aşağıdaki şekilde görülen karmaşıklık matrisi kullanılmaktadır (Tan ve ark., 2014). Bu çalışmada performans değerlendirmesi için denklem (1) ve (4) verilen değerler ve ROC eğrisi altında kalan AUC değeri kullanılmıştır (Fawcett, 2006). AUC değeri bire ne kadar yakınsa model o kadar başarılı olarak değerlendirilmektedir.

DP: Hasta olup, hasta olarak tahmin edilen

YN: Hasta olup, hasta değil olarak tahmin edilen

YP: Hasta olmayıp, hasta olarak tahmin edilen

DN: Hasta olmayıp, hasta değil olarak tahmin edilen

**Tablo 3.** Karmaşıklık matrisi

		Tahminlenen	
		Doğru	Yanlış
Gerçek	Doğru	DP	YN
	Yanlış	YP	DN

$$\text{Doğruluk Oranı} : \frac{(DP+DN)}{DP+YN+DN+YP} \quad (1)$$

$$\text{Hassasiyet} : \frac{DP}{DP+YP} \quad (2)$$

$$\text{Geri Çağırma} : \frac{DP}{DP+YN} \quad (3)$$

$$\text{F1 Skoru} : 2x \left( \frac{\text{Hassasiyet} \times \text{Geri Çağırma}}{\text{Hassasiyet} + \text{Geri Çağırma}} \right) \quad (4)$$

### 3. BULGULAR VE TARTIŞMA

Tablo 4’de veri setleri üzerinde herhangi bir veri ön işleme tekniği kullanılmadan ve modelin eğitilmesinde parametre optimizasyonu yapılmadan eğitilen Naive Bayes modellerin başarıları görülmektedir.



**Tablo 4.** Ham veri ile NB modellerinin performansları

NAIVE BAYES	ÇOCUK VERİ SETİ	ERGEN VERİ SETİ	YETİŞKİN VERİ SETİ
Doğruluk Oranı %	93.80	88.73	96.44
AUC	0.982	0.963	0.994
Hassasiyet %	89.99	86.43	92.98
F1 Skoru %	93.26	90.26	93.20

Daha sonra veri setlerinden aykırı veriler silinip, satırdaki eksik değerler o sütunun ortalama değeri alınıp, eksik değerler tamamlanmıştır.

**Tablo 5.** Yetişkin veri seti en başarılı model eğitme parametreleri

PARAMETRE	
Laplace Doğrulaması	Evet
Tek Çıkışlı Çapraz Doğrulama	Hayır
Öznitelik Eşik Ağırlığı	0.02
Çapraz Doğrulama Katlama Sayısı	8

**Tablo 6.** Veri ön işleme ve parametre optimizasyonu sonrası yetişkin veri seti ile NB modelinin performansı

NAIVE BAYES	YETİŞKİN VERİ SETİ
Doğruluk Oranı %	98.71
AUC	0.999
Hassasiyet %	97.33
F1 Skoru %	97.57

Birçok parametre değeri denenerek yetişkin veri seti ile NB modelleri oluşturulup performans değerleri kayıt edilmiştir. Tablo 5’ de görülen parametre değerleri ile en başarılı NB modeli oluşturulmuş ve Tablo 6’ da görüldüğü üzere veri ön işleme ve parametre optimizasyonu sayesinde yetişkin veri seti üzerinden daha başarılı NB modeli elde edilmiştir.

**Tablo 7.** Ergen veri seti en başarılı model eğitme parametreleri

PARAMETRE	
Laplace Doğrulaması	Hayır
Tek Çıkışlı Çapraz Doğrulama	Hayır
Çapraz Doğrulama Katlama Sayısı	2
Öznitelik Eşik Ağırlığı	0.1

**Tablo 8.** Veri ön işleme ve parametre optimizasyonu sonrası ergen veri seti ile NB modelinin performansı

PARAMETRE	
Doğruluk Oranı %	95.74
AUC	0.985
Hassasiyet	96.67
F1 Skoru	96.69

Aynı işlem ergen veri seti üzerinde de tekrarlanmıştır. Tablo 7’ de en başarılı model eğitilirken kullanılan parametreler görülmektedir. Tablo 8’ de modelin performans sonuçları incelendiğinde ham veri ile ve parametre optimizasyonu yapılmadan eğitilen modele göre çok daha başarılı model elde edilmiştir.

**Tablo 9.** Çocuk veri seti en başarılı model eğitme parametreleri

PARAMETRE	ÇOCUK VERİ SETİ
Laplace Doğrulaması	Evet
Çapraz Doğrulama Katlama Sayısı	25
Tek Çıkışlı Çapraz Doğrulama	Hayır
Öznitelik Eşik Ağırlığı	0.02

**Tablo 10.** Veri ön işleme ve parametre optimizasyonu sonrası çocuk veri seti ile NB modelinin performansı

NAIVE BAYES	ÇOCUK VERİ SETİ
Doğruluk Oranı %	97.58
Auc	0.994
Hassasiyet %	95.03
F1 Skoru %	97.03

Son olarak çocuk veri seti üzerinde veri ön işleme teknikleri sonrası modeller eğitilip test edilmiştir. En iyi başarı oranlarına sahip model eğitilirken Tablo 9’da görülen parametre değerleri kullanılmıştır. Bu parametreler ile oluşturulan modelin başarı oranları Tablo 10’ da görülmektedir. Veri ön işleme ve parametre optimizasyonu sonucunda modellerin başarısı yüzde 2 ile yüzde 8 arasında artış göstermiştir. Üç veri seti için de ağırlık hesaplaması yaparken gini indeksi ve korelasyon teknikleri kullanılmıştır. Gini indeksi ile belirlenen ağırlıklarla yapılan öznitelik seçiminde modeller %1 ve %2 arasında daha iyi başarı oranı sağlamıştır. En iyi başarı 98.71 ile yetişkin veri setinden oluşturulan modelden elde edilmiştir. En yüksek performans artışı ise yüzde 7.01 ile ergen veri seti üzerinden eğitilen modelden elde edilmiştir.

Literatürdeki çalışmalar incelendiğinde Naive bayes algoritması otizm teşhisinde denenmiş ve diğer algoritmalar ile karşılaştırılmıştır. NB algoritmasının da olduğu çalışmalarda NB yetişkinlik dönemi için %93 ve %96 arasında başarı, ergenlik dönemi için ortalama olarak %91 başarı ve çocukluk dönemi için yapılan çalışmalarda ise ortalama %95 ve %96 arasında doğruluk oranı göstermiştir ancak bu çalışma ile otizm teşhisinde NB algoritmasının başarısının arttırılabileceği görülmektedir. Şekil 1' de verilen yöntem ile NB algoritmasının literatürdeki çalışmalara göre yetişkinlik dönemi için ortalama %5, ergenlik dönemi için %4 ve çocukluk dönemi için ise %2 daha başarılı modeller oluşturduğu tespit edilmiştir. Bu çalışmadaki yöntem kullanılarak daha sonraki çalışmalarda diğer algoritmalarla denenip algoritmalar arasında daha doğru sonuç veren karşılaştırmalar yapılabilir.

#### 4. SONUÇLAR

Bu çalışmada üç ayrı veri seti kullanılarak Otistik Spektrum Bozukluğunun erken teşhisi için Naive Bayes algoritması ile sınıflandırma yapılmıştır. Literatürdeki çalışmalar incelendiğinde yapılan çalışmaların bir kısmında NB algoritmasının diğer algoritmalarla karşılaştırılmadığı, karşılaştırıldığında ise doğruluk oranına göre diğer algoritmaların gerisinde görüldüğü için ve NB en fazla öne çıkan makine öğrenmesi algoritmalarından biri olduğu için bu çalışmada NB algoritması, otizm teşhisinde değerlendirilmiştir.

Ham veri ve model eğitilirken herhangi bir parametre optimizasyonu yapılmadan elde edilen başarı değerleri ile aykırı veriyi temizleme, eksik değerleri ortalama değer ile doldurma ve öznitelik seçimi, çapraz doğrulama katlama sayısı gibi parametrelerde optimizasyonlar yapılarak modellerin başarıları karşılaştırılmıştır. Veri ön işleme teknikleri ve parametre optimizasyonu yapıldığında Naive Bayes algoritmasının üç veri seti üzerinde de başarı oranlarının arttığı görülmüştür.

#### KAYNAKÇA

- Abdullah, A. A., Rijal, S., Dash, S. R. (2019). Evaluation on Machine Learning Algorithms for Classification of Autism Spectrum Disorder (ASD). *In Journal of Physics: Conference Series* 1372 1 012052.
- Alwidian, J., Elhassan, A., Ghnemat, R. (2020). Predicting autism spectrum disorder using machine learning technique. *International Journal of Recent Technology and Engineering*, 8(5), 4139-4143.
- Atan, S. (2020). KNN, Naive Bayes ve Karar Ağacı Makine Öğrenme Algoritmaları, Bu Algoritmaların Sosyal Bilimlerde Kullanım İmkânları. *SocArXiv*. doi:10.31235/osf.io/8r5pu

- Büyükoflaz, F. N. ve Öztürk, A. (2018). "Early autism diagnosis of children with machine learning algorithms,". *Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4, doi: 10.1109/SIU.2018.8404223.
- Cho S, Liberman M, Ryant N, Cola M, Schultz RT, Parish-Morris J. (2019). Automatic Detection of Autism Spectrum Disorder in Children Using Acoustic and Text Features from Brief Natural Conversations. *Proceedings of the Annual Conference of INTERSPEECH*; (pp. 2513-2517), Graz, Austria.
- Demirhan, A. (2018). Performance of machine learning methods in determining the autism spectrum disorder cases. *Mugla Journal of Science and Technology*, 4(1), 79-84.
- Devika Varshini, G., Chinnaiyan, R. (2020). Optimized Machine Learning Classification Approaches for Prediction of Autism Spectrum Disorder. *Ann Autism Dev Disord*, 1(1), 1001.
- Anonim, (2023a) "Autistic Spectrum Disorder Screening Data for Adolescent", <https://archive.ics.uci.edu/ml/machine-learning-databases/00420/>. (Erişim Tarihi: 12.02.2023).
- Anonim, (2023b) "Autistic Spectrum Disorder Screening Data for Adult", <https://archive.ics.uci.edu/ml/machine-learning-databases/00426/>. (Erişim Tarihi: 12.02.2023).
- Anonim, (2023c) "Autistic Spectrum Disorder Screening Data for children", <https://archive.ics.uci.edu/ml/machine-learning-databases/00419/>. (Erişim Tarihi: 12.02.2023).
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861-874.
- Jiang, L., Zhang, L., Li, C., Wu, J. (2018). A correlation-based feature weighting filter for naive Bayes. *IEEE transactions on knowledge and data engineering*, 31(2), 201-213.
- Kotsiantis, S. B., Kanellopoulos, D., Pintelas, P. E. (2006). Data preprocessing for supervised learning. *International journal of computer science*, 1(2), 111-117.
- Manek, A. S., Shenoy, P. D., Mohan, M. C. (2017). Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. *World wide web*, 20(2), 135-154.
- Metlek, S., & Kayaalp, K. (2020). Otistik Spektrum Bozukluğunun Makine Öğrenme Algoritmaları ile Tespiti. *Zeki Sistemler Teori ve Uygulamaları Dergisi*, 3(2), 60-68.
- Mythili, M. S., & Shanavas, A. M. (2014). A study on Autism spectrum disorders using classification techniques. *International Journal of Soft Computing and Engineering*, 4(5), 88-91.
- Özkan, Ş. Y., Ergenekon, Y., Çolak, A., Kaya, Ö., Cavkaytar, S. (2015) Otizm spektrum bozukluğu. A. Cavkaytar (Ed.), 22-49, *Aile ve Sosyal Politikalar Bakanlığı*, Ankara.
- Parikh, M. N., Li, H., He, L. (2019). Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data. *Frontiers in computational neuroscience*, 13, 9.
- Raj, S., & Masood, S. (2020). Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques. *Procedia Computer Science*, 167, 994-1004.

Tan, P., Steinbach, M., Kumar, V. (2014). “Performance Measure” in Introduction to Data Mining, Pearson Education Limited (UK).

Su, M. & Takcı H. (2023). Naive Bayes ile Otistik Spektrum Bozukluđu Tanısı Koyma. *Şırnak Üniversitesi Fen Bilimleri Dergisi*, 3(2), 50-62.

Su, M. & Takcı H. (2023). Diagnosing Autistic Spectrum Disorder with Naïve Bayes. *Sirnak University Journal Of Science*, 3(2), 50-62.