Fırat Üniversitesi Müh. Bil. Dergisi
37(1), 453-466, 2025

Araştırma Makalesi
https://doi.org/10.35234/fumbd.1607609

TÜİK Mikro Verileri ile Çocuk İşgücü Tahmini: Makine Öğrenimi Modellerinin Performans Analizleri

Kamil Abdullah EŞİDİR[1*]
[1] Fırat Kalkınma Ajansı, Elazığ, Türkiye
[*1] abdullahesidir@yahoo.com

(Geliş/Received: 26/12/2024; Kabul/Accepted: 03/02/2025)

Öz: Çalışma, Türkiye İstatistik Kurumu (TÜİK) tarafından 2019 yılında gerçekleştirilen Çocuk İşgücü Araştırması mikro veri setini kullanarak, çocuk işçiliği konusunu makine öğrenimi modelleri ile analiz etmektedir. 5-17 yaş grubundaki çocukların çalışma durumu, eğitim faaliyetleri ve ailevi sorumlulukları gibi faktörler incelenerek çocuk işçiliğini etkileyen unsurlar belirlenmiştir. Yaş grubu, çocukların çalışma durumunu en fazla etkileyen bağımsız değişken olup, erkek çocukların çalışma oranı kızlara göre daha yüksektir. Eğitimine devam etmeyen çocukların çalışma olasılığı ise artış göstermektedir. Çocukların çalışma durumu; Random Forest, XGBoost, Gradient Boosting ve SVM makine öğrenimi modelleri kullanılarak tahmin edilmiştir. Modellerin performansları; doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 Skoru metrikleri ile değerlendirilmiştir. Random Forest ve SVM modelleri, F1 Skoru açısından en yüksek performansı göstermiştir. Elde edilen bulgular, çocuk işçiliğini etkileyen faktörlerin makine öğrenmesi modelleri ile derinlemesine anlaşılmasına katkıda bulunmakta ve uygulanacak politikaların veri temelli şekillendirilmesinin önemini vurgulamaktadır.

Anahtar kelimeler: Makine Öğrenimi, Çocuk İşgücü Araştırması, Yönetim Bilişim Sistemleri, Mikro Veri, XGBoost.

Child Labor Forecasting with TurkStat Micro Data: Performance Analysis of Machine Learning Models

Abstract: The study analyzes the issue of child labor with machine learning models using the micro dataset of the Child Labor Force Survey conducted by the Turkish Statistical Institute (TurkStat) in 2019. Factors affecting child labor are identified by examining factors such as the working status, educational activities and family responsibilities of children in the 5-17 age group. Age group is the independent variable that affects the working status of children the most, and boys are more likely to work than girls. Children who do not continue their education are more likely to work. The working status of children is estimated using Random Forest, XGBoost, Gradient Boosting and SVM machine learning models. The performance of the models was evaluated with accuracy, precision, recall and F1 Score metrics. Random Forest and SVM models showed the highest performance in terms of F1 Score. The findings contribute to an in-depth understanding of the factors affecting child labor with machine learning models and emphasize the importance of shaping policies based on data.

Key words: Machine Learning, Child Labor Survey, Management Information Systems, Microdata, XGBoost.

1. Giriş

Çocuk işçiliği; küresel çapta ekonomik, sosyal ve kültürel çeşitli faktörlerin etkisiyle ortaya çıkan önemli bir toplumsal sorun olarak tanımlanmaktadır. Uluslararası Çalışma Örgütüne (ILO) göre, 5-17 yaş arasındaki fertleri kapsamaktadır. Çocuk işçiliği; küresel anlamda önemli bir toplumsal sorun olup, ekonomik, sosyal ve kültürel faktörlerin bir sonucu olarak ortaya çıkmaktadır [1]. Uluslararası Çalışma Örgütü'nün (ILO) önerileri doğrultusunda, çocuk işçiliğiyle mücadele kapsamında gerçekleştirilen araştırmalar, çocukların çalışma koşullarını, eğitim durumlarını ve sosyal yaşantılarını detaylı bir şekilde analiz etmektedir. Çalışmada kullanılan mikro veri seti, Türkiye İstatistik Kurumu (TÜİK) tarafından 2019 yılında gerçekleştirilen Çocuk İşgücü Araştırmasından (ÇİA) elde edilmiştir [2]. Araştırma 5-17 yaş grubundaki çocukların istihdam durumlarını, eğitim faaliyetlerini ve ailevi durumlarını inceleyerek, çocuk işçiliğinin mevcut durumunu değerlendirme ve çözüm önerileri geliştirme amacı gütmektedir.

Çalışmada, çocuk işgücüne dair cinsiyet, yaş grubu, eğitime devam durumu, en son bitirilen okul, hane halkı sorumlusunun çalışma durumu ve çocukların istihdam durumu değişkenleri ele alınmış ve makine öğrenmesi modelleri kullanılarak analizler ve tahminlemeler yapılmıştır. Aynı zamanda, bağımsız değişkenlerin model performansları üzerindeki etkileri analiz edilerek, çocuk işçiliğinin önlenmesine yönelik politika önerilerinde

* Sorumlu yazar: abdullahesidir@yahoo.com Yazarın ORCID Numarası: [1] 0000-0002-8106-1758

bulunulmuştur. Çalışma, akademik literatüre katkı sağlarken, politika yapıcılar için somut veri tabanlı bulgular sunmaktadır. Ekonomik analizlerde ve yapılan tahminlemelerde, gelişmiş veri analitiği model ve yöntemleri giderek daha önemli hale gelmektedir. Geliştirilen modeller, büyük veri setlerini işlemekte ve karmaşık ilişkiler ile eğilimleri anlayarak, analizcilere kapsamlı, yorumlanabilir ve değerli bilgiler sunmaktadır [3]. Makine öğrenimi modellerinin geleneksel modellere kıyasla daha geniş bir değişken yelpazesi ile çalışabilmesi ve karmaşık etkileşimleri modelleyebilmesi, çalışmada tercih edilmesine neden olmuştur. RandomForest, XGBoost ve Gradient Boosting ve benzeri modeller, özellikle büyük veri setlerinde yüksek doğruluk ve güvenilir sonuçlar elde edilmesine olanak sağlayarak, analizlerin etkinliğini artırmaktadır.

Geleneksel istatistiksel yöntemler, regresyon analizleri, lojistik regresyon ve ANOVA gibi teknikler, belirli koşullar altında ve sınırlı değişkenler ile etkili olabilmektedir [4]. Büyük veri setleri ve çok boyutlu veri yapıları söz konusu olduğunda, geleneksel yöntemler ve modeller yetersiz kalmakta ve karmaşık ilişkileri açığa çıkaramamaktadır [5]. Buna karşılık, makine öğrenmesi modelleri ise, büyük ve karmaşık veri kümelerinden öğrenme yeteneği sayesinde, geleneksel yöntemler ile aşılamayan birçok zorluğu kolaylıkla aşabilmektedir [6]. Makine öğrenimi modellerinin tercih edilmesinin temel nedeni, verinin karmaşıklığını ve çeşitliliğini daha iyi yönetebilme yeteneğidir [7].

2. Literatür Taraması

Yapılan çalışmalar, geleneksel istatistiksel yöntemlerin kapsamlı veri kümelerindeki karmaşık ilişkileri ve gizli kalıpları ortaya çıkarmada yetersiz olduğunu ve daha gelişmiş teknik ve modellerin kullanılması gerekliliğini göstermiştir [8]. Makine öğrenimi model ve algoritmaları kullanılarak, daha geniş bir değişken yelpazesini dikkate alan daha doğru tahmin modelleri geliştirilebilmektedir. Makine öğrenimi algoritmaları, daha geniş bir değişkenler dizisini ve bunların etkileşimlerini dikkate almakta ve çok daha kapsamlı yaklaşımlar sunmaktadır [9].

Çöpoğlu'nun [10] "Türkiye'de Çocuk İşçiliği" çalışması, çocuk işçiliğinin ekonomik, sosyal ve kültürel nedenlerini kapsamlı bir şekilde ele alarak, bu durumun birey ve toplum üzerindeki olumsuz etkilerini ortaya koymaktadır. Çocuk işçiliği; yoksulluk, eğitime erişim kısıtlılıkları, göç ve kültürel faktörler ile ilişkilendirilmiştir. Çalışmada, özellikle eğitimden yoksun kalan çocukların fiziksel ve psikolojik gelişimlerinin zarar gördüğü vurgulanmıştır. Çalışma; çocuk işçiliğinin azaltılması için eğitim imkanlarının artırılması, yasal düzenlemelerin etkinleştirilmesi ve sosyal politikaların gözden geçirilip güçlendirilmesi gerekliliğini savunmaktadır.

Gül ve Öztürk [11] çalışmalarında; çocuk işçiliğinin ekonomik, sosyal ve kültürel nedenlerini kavramsal çerçevede incelemiş, bu durumun çocukların fiziksel ve zihinsel gelişimlerine zarar verirken toplumsal sorunları da derinleştirdiğini vurgulamıştır. Çalışmada, yoksulluk, eğitimsizlik ve göçün çocuk işçiliğini tetikleyen başlıca faktörler olduğu belirtilmiştir. Ayrıca, çocukların eğitimden uzaklaşmalarının kısır bir döngü meydana getirdiği; eğitimsizlik, işsizlik ve yoksulluğun birbirini besleyerek toplumun refahını olumsuz etkilediği anlatılmıştır. Çocuk işçiliğiyle mücadelede yoksulluğu azaltıcı politikalar, nitelikli eğitim fırsatları sunulması ve sosyal koruma mekanizmalarının geliştirilmesi önerilmiştir.

Tosunoğlu ve arkadaşları [12], makine öğrenmesi modellerinin eğitim bilimlerindeki uygulamalarını incelemiştir. 2015-2020 yılları arasında yayımlanan 184 makale üzerinde yapılan içerik analizi sonucunda, en sık kullanılan algoritmaların karar ağaçları (%22,5), destek vektör makineleri (%17,8) ve Naive Bayes (%14,2) olduğu belirlenmiştir. Çalışmaların çoğu deneysel veya uygulamalı olup, lisans öğrencileri ile ve veri tabanlarından elde edilen verilerle gerçekleştirilmiştir. Yapılan çalışma, eğitimde makine öğrenmesi kullanımına dair eğilimleri ortaya koyarak literatüre önemli katkılar sunmaktadır.

Eriş Dereli [13], TÜİK'in 2019 yılı Çocuk İşgücü Araştırması mikro verilerini kullanarak, Türkiye'deki çocuk istihdamını etkileyen faktörleri analiz etmiştir. Çalışma; kız çocuklarının çalışma olasılığının erkek çocuklara göre daha düşük olduğunu, eğitimine devam eden çocukların istihdam olasılığının azaldığını ve anne-baba eğitim seviyesinin artmasının çocuk işçiliğini azalttığını ortaya koymuştur. Ayrıca, çocuk yaşının artması ve ebeveynlerin işsiz olmasının istihdam olasılığını artırdığı tespit edilmiştir. Çalışma, çocuk işçiliği ile mücadelede kadınların eğitimi ve istihdam politikalarına yönelik öneriler ile literatüre önemli katkılar sunmaktadır.

Kömüryakan ve Çağlayan [14] çalışmalarında, Türkiye'deki çocuk işçiliğini etkileyen sosyal, ekonomik ve demografik faktörleri; simetrik (Logit ve Probit) ve asimetrik (Log-log ve Clog-log) modelleri kullanarak analiz etmişlerdir. Çalışma; Türkiye'nin çocuk işçiliği oranının dünya ortalamasının altında olmasına rağmen, Avrupa ve Kuzey Amerika ülkelerine kıyasla iki kat daha fazla olduğunu göstermiştir. Çalışmada; çocuk işçiliğinin özellikle 15-17 yaş aralığındaki erkek çocuklar ile küçük kardeşi bulunan ve kalabalık hanelerde yaşayan bireylerde yoğunlaştığı bulunmuştur. Elde edilen bulgulara göre, çocuk işçiliğini önlemeye yönelik politikalarda hedef gruplara odaklanmanın önemi vurgulanmaktadır.

Yardımcıoğlu [15], çocuk işçiliğiyle mücadelede Avrupa Birliği'nin "Kurumsal Sürdürülebilirlik Durum Tespitine Dair Direktif Önerisi"ni detaylı şekilde ele almıştır. Çalışma, çocuk işçiliğinin fiziksel ve zihinsel gelişim üzerindeki olumsuz etkilerini vurgulamaktadır. Şirketlere tedarik zincirleri boyunca insan hakları ihlallerini önleme ve raporlama yükümlülükleri getirilmektedir. Çocuk işçiliğini engellemek adına denetim mekanizmaları oluşturması ve cezai ve idari sorumluluklar üstlenilmesi öngörülmektedir. Yapılan düzenlemenin, çocuk işçiliği ile mücadelede küresel ölçekte önemli bir hukuki adım olduğu ve Türkiye dahil birçok ülkedeki politikalar için yol gösterici olduğu anlatılmaktadır.

Özdemir ve arkadaşları [16], Türkiye'de çocuk işçiliğinin sağlık ve sosyal yaşam koşulları üzerindeki etkilerini incelemiştir. Çalışma; çocuk işçiliğinin fiziksel, zihinsel ve sosyal gelişimi olumsuz yönde etkilediğini ve bu durumun eğitimden kopuş, yoksulluk döngüsünün devamı gibi sonuçlara yol açtığını ortaya koymuştur. Çalışmada, çocukların çalışma koşullarının iyileştirilmesi ve çocuk işçiliğinin önlenmesi adına sosyal politikaların etkin biçimde uygulanmasının önemi vurgulanmıştır. Çocuk refahının artırılmasına yönelik yapılacak çalışmalarda kapsamlı veri temelli yaklaşımların geliştirilmesi gerektiği ifade edilmiştir.

3. Metodoloji

Çalışmada kullanılan metodoloji, sınıflandırma performansını iyileştirmek için farklı makine öğrenimi modellerini göstermeyi amaçlamaktadır. Analizlerde, çeşitli veri ön işleme teknikleri, model eğitim süreçleri, özellik çıkarımı yöntemleri kullanılmıştır. Büyük veri, çeşitlilik, hız ve hacimden ötürü çeşitli zorluklara yol açmaktadır [17]. Makine öğrenmesi veriye dayalı problemlerin çözümünde, uygun algoritmalar ile modelleme yapmayı hedefleyen bir çeşit hesaplama disiplinidir [18].

3.1. Yazılım ve Donanım

Veri analizi ve makine öğrenimi süreçlerinin yüksek verimlilik ile gerçekleştirilebilmesi için çalışmada güçlü bir donanım kullanılmıştır. Windows 11 Pro işletim sistemiyle çalışan 64-bit Intel64 Family 6 Model 142 Stepping 10 işlemciye sahip bir bilgisayar tercih edilmiştir. Cihaz, geniş veri kümelerinin işlenmesi ve modellerin hızlı eğitimi için yeterli olan 16 GB RAM barındırmaktadır. Analiz süreçlerinde Python programlama dili kullanılmıştır. En güncel sürüm olan Python 3.12.7 tercih edilerek, veri analizi ve bilimsel hesaplama süreçlerinde geniş bir kütüphane desteğinden (NumPy, Pandas, Scikit-learn, TensorFlow, Keras vs.) faydalanılmıştır. Çalışmaların daha verimli ve anlaşılır bir şekilde yürütülebilmesi için platform olarak, Jupyter Notebook 7.2.2 sürümü tercih edilmiştir. Bu platform, kodlama, veri görselleştirme ve açıklamaların entegre bir yapıda sunulmasını sağlayarak analiz süreçlerini kolaylaştırmaktadır.

3.2. Veri Seti

Verilerin resmi kaynaklardan elde edilmesi, diğer araştırmacıların benzer analizler yapmasına olanak tanımakta, aynı zamanda güvenilir ve kamuya açık veri kaynaklarının kullanımını da teşvik etmektedir [19]. Çalışmada, orijinal mikro veri setinde yer alan eksik veriler temizlenmiş ve analizlere uygun hale getirilmiştir. 25.190 gözlem içeren veri setinde, biri bağımlı ve 5'i bağımsız olmak üzere toplamda 6 değişken bulunmaktadır. Çocuğun çalışıp çalışmadığı durumunu ifade eden COCUK_CALISAN değişkeni, bağımlı değişken olarak seçilmiş ve diğer değişkenler bağımsız değişkenler olarak analizlerde kullanılmıştır. Veri setinde yer alan değişkenlerin isimleri, açıklamaları ve kodlama seçenekleri Tablo 1'de sunulmuştur. Veri setinde bulunan değişkenler; 5-17 yaş arasındaki çocukların cinsiyet, yaş grubu, eğitim durumu ve istihdam gibi demografik ve sosyal özelliklerini belirtmektedir.

Tablo 1. Çalışmada kullanılan veri seti, değişkenler ve açıklamaları.

Değişken Adı	**Açıklama**	**Kodlama Seçenekleri**
CINSIYET	Cinsiyet	1: Erkek 2: Kadın
YAS_GRUP	Yaş grubu (5-17 yaş arası)	1: 5-11 yaş 2: 12-14 yaş 3: 15-17 yaş
OKUL_BITEN	En son bitirilen okul	0: Bir okul bitirmeyen 1: İlkokul 2.1: Genel ortaokul 3.1: Genel lise 3.2: Mesleki veya teknik lise
EGITIM_DEVAM	Eğitime devam durumu	1: Eğitime devam eden 2: Eğitime devam etmeyen
HHS_DURUM	Hanehalkı sorumlusunun çalışma durumu	1: İstihdamda 2: İşsiz 3: İşgücüne dahil olmayan
COCUK_CALISAN	5-17 yaş grubundaki istihdam durumu	0: İstihdamda 1: İstihdamda değil

3.2. Veri Ön İşleme

Veri ön işleme, veri setinin analiz ve modelleme için hazır hale getirilmesi sürecidir. Bu aşamada gerçekleştirilen işlemler, modellerin doğruluğunu ve genel performansını önemli ölçüde etkilemektedir [20]. Modellerin eğitilmesi esnasında, ihtiyaç duyulmayan ve çok fazla eksik değerler içeren sütunlar, veri setinden çıkarılmıştır. Hedef (bağımlı) değişken, "İstihdamda" değeri için 0 ve "İstihdamda değil" değeri için 1 olacak şekilde ikili formata dönüştürülmüştür. Kategorik değişkenler, makine öğrenmesi algoritmalarıyla uyumlu hale getirilmek için sayısal değerlere dönüştürülmüştür. Ardından, dönüştürülen veri seti, eğitim (%80) ve test (%20) setlerine ayrılmıştır. Sayısal veriler, Standard Scaler kullanılarak ölçeklendirilmiştir. Bu işlem, özellikler arasındaki ölçek farklılıklarını gidererek modelin daha etkili öğrenmesine yardımcı olmaktadır [21]. Matematiksel olarak bu işlem Denklem 1'de gösterildiği şekilde ifade edilmektedir:

$$z = \frac{(x-\mu)}{\sigma} \tag{1}$$

Burada, x ölçeklendirilecek olan özellik değerini, μ bu özelliğin ortalamasını, σ bu özelliğin standart sapmasını ve z ölçeklendirilmiş sonucu ifade etmektedir.

3.3. Modellerin Performans Değerlendirme Kriterleri

Makine öğrenmesinde analizi gerçekleştirilen modellerin performansını objektif bir şekilde değerlendirmek için çeşitli metrikler ve yöntemler kullanılmaktadır [22]. Çalışmada, makine öğrenmesi modellerinin performansları, Kesinlik (Precision), Duyarlılık (Recall), F1 Skoru ve Doğruluk (Accuracy) metrikleri ile ölçülmüştür. Bu metriklerin, matematiksel formülleri sırasıyla Denklem 2, 3, 4 ve 5'te ifade edilmiştir. Denklemlerde, DP doğru pozitif, YP yanlış pozitif, DN doğru negatif ve YN yanlış negatiftir. Ayrıca, tahminlerin doğruluğunu görsel şekilde değerlendirebilmek için karmaşıklık matrisleri de kullanılmıştır.

$$\text{Kesinlik (Precision)} = \frac{\text{DP}}{\text{DP+YP}} \tag{2}$$

$$\text{Duyarlılık (Recall)} = \frac{\text{DP}}{\text{DP+YN}} \tag{3}$$

$$\text{F1 Skoru} = 2 \times \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik+Duyarlılık}} \tag{4}$$

$$\text{Doğruluk (Accuracy)} = \frac{\text{DP+DN}}{\text{DP+DN+YP+YN}} \quad (5)$$

Makine öğrenmesi modellerinin performansını değerlendirmek için kullanılan Kesinlik, Duyarlılık, F1 Skoru ve Doğruluk metrikleri, farklı senaryolarda modelin gücünü ve zayıflıklarını belirtmesi açısından önemlidir [22].

4. Makine Öğrenmesi Modelleri

Makine öğrenimi, bilgisayar kullanımı ile verilerin analiz edilerek öğrenilmesi ve tahmin yapma yeteneği kazandırmayı hedefleyen bir bilim dalıdır. Çözüm süreçlerinde farklı matematiksel ve istatistiksel modeller kullanılmakla birlikte, modellerin seçimi veri kümesinin türüne ve özelliklerine göre belirlenmektedir.

4.1. XGBoost (Extreme Gradient Boosting)

XGBoost, Gradient Boosting algoritmalarının geliştirilmiş bir versiyonudur ve yüksek performans gösteren modellerden birisidir. Makine öğrenmesinde sınıflandırma ve regresyon problemlerinde sıkça kullanılmaktadır. Hızlı ve yüksek doğruluk oranları sunması, onu veri bilimi yarışmaları ve büyük ölçekli projelerde popüler hale getirmiştir. XGBoost, son yıllarda makine öğrenmesi alanında en başarılı ve yaygın kullanılan yöntemlerden birisidir. Gereksiz e-postaların sınıflandırılması (Spam), reklam eşleştirme, dolandırıcılık tespiti ve fiziksel olaylardaki anomali tespiti gibi birçok alanda etkili sonuçlar sunmaktadır [23]. XGBoost, doğru tahminler üretmede etkili olduğunu ispatlamıştır ve birçok alternatif makine öğrenimi yaklaşımına göre hesaplama avantajları sunmaktadır [24]. Özellikle büyük veri setlerinde ve karmaşık problemlerde sağladığı etkili performans, XGBoost'u popüler bir tercih haline getirmektedir.

XGBoost, gradyan artırma algoritmasına dayalı olarak çalışan bir topluluk öğrenme yöntemidir. Bu teknik, bir hata fonksiyonunu en aza indirmek için bir dizi zayıf öğreniciyi sırayla birleştirmeyi ve böylece tahmin doğruluğunu artırmayı içerir [23]. XGBoost, ölçeklenebilirliği ve esnekliği nedeniyle popülerlik kazanmış ve gradyan artırma kütüphanesi içinde giderek yükselen bir topluluk öğrenme yöntemi haline gelmiştir. Torbalama ve istifleme gibi yaygın olarak kullanılan diğer topluluk yöntemlerinin yanı sıra, boosting algoritması kategorisine giren bir topluluk öğrenme algoritması olarak kabul edilir [25]. Extreme Gradient Boosting'in kısaltması olan XGBoost, gradient boosting ailesinin bir parçasıdır ve zayıf öğrenenleri güçlü öğrenenlere dönüştürmeyi amaçlar. Bir topluluk öğrenme algoritması olarak XGBoost, boosting yaklaşımını kullanarak birden fazla karar ağacını sağlam bir sınıflandırıcıya entegre etmektedir [26]. Modelin tahmin fonksiyonu Denklem 6'da gösterilmektedir:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in F \quad (6)$$

XGBoost, her iterasyonda hatayı minimize etmek için gradyan iniş algoritmasını kullanır ve bu sayede modelin doğruluğunu artırır. Modelin genel kayıp fonksiyonu Denklem 7'deki gibidir:

$$L(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (7)$$

4.2. RandomForest (Rastgele Orman)

RandomForest, farklı alt veri setlerinde birden fazla karar ağacının birleşiminden oluşmaktadır. Random Forest, sınıflandırma ve regresyon görevleri için kullanılan, topluluk öğrenme metoduna dayanan bir makine öğrenimi algoritmasıdır. Bu algoritma, birden fazla karar ağacını bir araya getirerek her bir ağacın zayıflıklarını dengelemekte ve daha güçlü ve dengeli bir model oluşturmaktadır [27]. Random Forest, eğitim veri setinden birçok bootstrap örneği çekmektedir (her örnek, orijinal veri setinin rastgele seçilmiş örneklerini içermektedir). Her bir bootstrap örneği için bağımsız bir karar ağacı kurulur. Her ağaç, veri setinin farklı bir alt kümesi üzerinde eğitilir. Her düğümdeki bölünme işlemi, tüm özellikler yerine rastgele seçilen bir alt küme kullanılarak yapılır. Bu, modelin varyansını azaltmaya yardımcı olur ve ağaçlar arası korelasyonu düşürür. Sınıflandırma için tahmin yapılırken genelde Denklem 8'de gösterildiği gibi bir ifade kullanılmaktadır:

$$y = \text{mod}\{y_1, y_2, \ldots, y_n\} \quad (8)$$

Burada, $y_1, y_2, \ldots, y_n$ her bir karar ağacının tahmin ettiği sınıf etiketleridir ve mod fonksiyonu en sık rastlanan etiketi belirlemektedir.

4.3. CatBoost (Category Boosting)

CatBoost, makine öğreniminde özellikle kategorik verilerle çalışmak için optimize edilmiş, güçlü bir gradyan artırma algoritmasıdır. İsmi "Category Boosting" ifadesinden kısaltılmış olup, diğer gradyan artırma algoritmalarına göre kategorik değişkenleri otomatik olarak işlemektedir. Bu özelliği, kategorik verilerin birden fazla dönüştürme işlemine tabi tutulmasını gerektirmediği için kullanıcı dostu bir deneyim sunar. CatBoost, sınıflandırma ve regresyon problemlerinde yüksek doğruluk oranları ile öne çıkmaktadır. Model, overfitting (aşırı öğrenme) riskini azaltmak için çeşitli düzenleme tekniklerini içerir ve veri dengesizliklerinden kaynaklanabilecek hataları minimize etmeyi amaçlar. Ayrıca, paralel hesaplama kapasitesi sayesinde, büyük veri kümelerinde hızlı eğitim süreçleri sunmaktadır.

CatBoost'un bir diğer avantajı ise, yerleşik GPU desteği sayesinde eğitim süreçlerini hızlandırabilmesi ve daha geniş veri setlerinde etkili performans sağlayabilmesidir. Model hiperparametreleri kolayca optimize edebilmektedir, bu durum da modelin farklı veri setlerine uyarlanabilmesini sağlamaktadır. CatBoost, Yandex firması tarafından geliştirilmiştir ve özellikle kategorik verileri etkili bir şekilde işleyebilmektedir. Geliştirilen model, kategorik özelliklerin doğrudan işlenmesini sağlayarak ön işleme ihtiyacını azaltmakta ve modelin doğruluğunu artırmaktadır [28]. CatBoostun tahmin doğruluğu, özellikle kategorik değişkenlerin yoğun olduğu veri setlerinde belirgin bir şekilde gözlemlenmektedir.

4.4. Destek Vektör Makineleri (SVM)

Destek Vektör Makineleri (SVM), hem sınıflandırma hem de regresyon görevlerinde kullanılan, güçlü ve esnek bir makine öğrenimi algoritmasıdır. Bu yöntem, sınıfları en iyi şekilde ayıran bir hiper düzlem bulmaya odaklanır ve doğrusal olarak ayrılabilir verilerde oldukça etkili sonuçlar verir. Doğrusal olarak ayrılamayan veri setlerinde ise kernel fonksiyonları kullanılarak yüksek boyutlu uzaylara geçiş sağlanır ve sınıflandırma doğruluğu artırılır. SVM, metin ve görüntü tanıma, biyometrik tanımlama, finansal tahminler ve biyoinformatik gibi çeşitli alanlarda başarıyla uygulanmaktadır. Özellikle, sınıflandırma problemlerinin çözümünde etkinliği kanıtlanmıştır [29].

Destek Vektör Regresyonu (DVR), Destek Vektör Makineleri (DVM) prensiplerini regresyon problemlerine genişleten bir makine öğrenme tekniğidir. DVR, tüm örnek noktalarının hiper düzlemden toplam sapmasını en aza indiren optimal bir hiper düzlem bulmayı amaçlamaktadır [30]. DVR'de algoritma, eğitim örneklerine dayalı olarak optimum bir regresyon hiper düzlemi oluşturur ve noktaları hiper düzleme yakın tutarak ayrımı en aza indirmeye çalışır [31]. SVR modeli, hiper düzlem ile eğitim verileri arasındaki marjı en üst düzeye çıkarmak için ikinci dereceden bir optimizasyon problemini çözer ve noktaları belirli bir hata toleransı dahilinde etkili bir şekilde kapsar. DVR, giriş verilerini doğrusal olmayan işlevler aracılığıyla düşük boyutlu bir uzaydan yüksek boyutlu bir özellik uzayına eşleyerek ve ardından örnek noktaların bu hiper düzleme mümkün olduğunca yakın düşmesini sağlamak için yüksek boyutlu özellik uzayında optimum bir hiper düzlem arayarak çalışır [32]. Bu süreç, eğitilmiş veri kümesine minimum mesafeyi sağlayan hiper düzlemi bulmayı ve veri örneklerine iyi uyması için hiper düzlemi optimize etmeyi içerir. DVR'nin prensibi, hiper düzlemden en uzaktaki örneklerin geometrik mesafesini en aza indiren ve veri dağılımını etkili bir şekilde yakalayan optimum hiper düzlemi aramaktır. Modelin matematiksel ifadesi Denklem 9'da gösterilmiştir.

$$y = \sum_{i=1}^{n}(a_i - a_i^*)K(x_i, x) + b \quad (9)$$

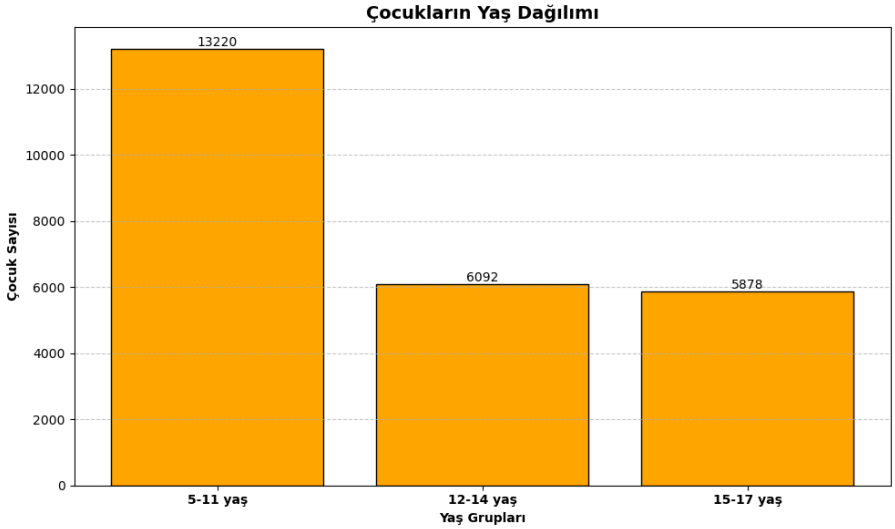
Burada, y tahmin edilen değeri, a_i ve a_i^* Lagrange çarpanlarını, K(x_i,x) çekirdek fonksiyonunu, b bias terimini ve x_i eğitim verisini temsil etmektedir. SVR, doğrusal olmayan ilişkileri modellemek için çekirdek fonksiyonları (lineer, polinomial, radyal bazlı fonksiyonlar gibi) kullanır. Çekirdek fonksiyonu K(x_i,x), yüksek boyutlu uzayda veri noktaları arasındaki benzerlikleri ölçmektedir.

5. Bulgular

Çalışmada, TÜİK tarafından 2019 yılında gerçekleştirilen Çocuk İşgücü Araştırması verileri kullanılarak, 5-17 yaş grubundaki çocukların istihdam durumları, eğitim süreçleri ve ailevi sorumluluklarına dair kapsamlı analizler yapılmıştır. Çalışmanın temel amacı, çocuk işçiliğini etkileyen faktörleri tespit ederek bu alanda etkin

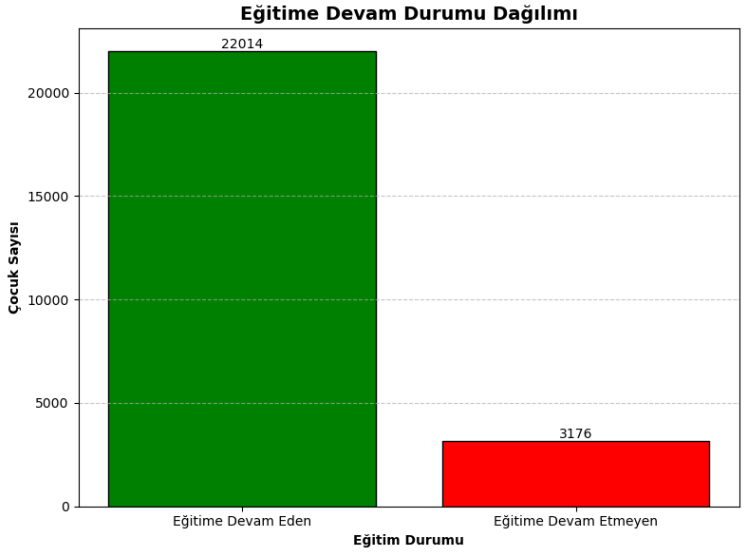
sosyal politika önerileri sunmaktır. Çocukların yaş, cinsiyet, eğitim durumu ve ailevi koşullar gibi demografik özellikleri, makine öğrenimi modelleri ile analiz edilerek çocuk işçiliğinin dinamiklerini daha iyi anlamak hedeflenmiştir.

Makine öğrenimi modellerinin kullanımı, geleneksel modellere kıyasla daha geniş değişken yelpazesini analiz etme ve karmaşık ilişkileri modelleme imkânı sunmaktadır [33]. Kullanılan Random Forest, SVM, XGBoost ve CatBoost modelleri; doğruluk, kesinlik, duyarlılık ve F1 skoru metrikleri ile değerlendirilmiştir. Analizler sonucu elde edilen bulgular, çocuk işçiliği ile mücadelede veri temelli politika önerilerinin önemini ortaya çıkartmıştır.



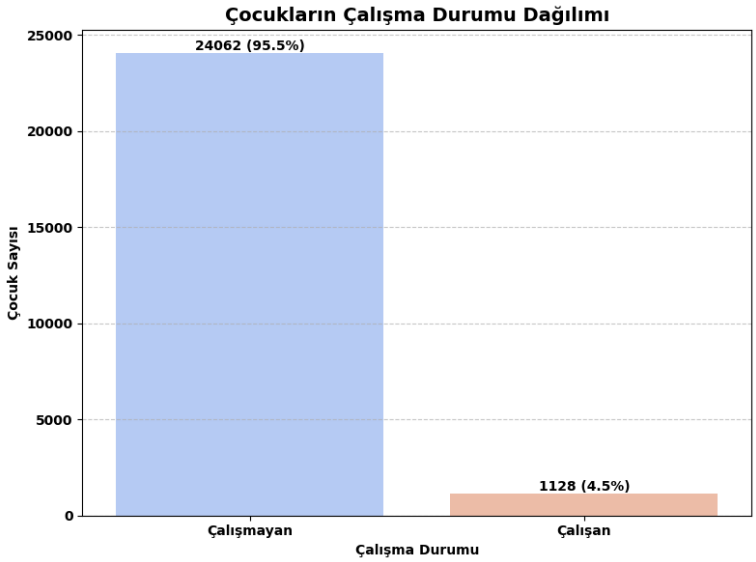
Şekil 1. 0-17 Yaş grubundaki çocukların yaş dağılımı.

Şekil 1'de 0-17 yaş grubundaki çocukların yaş dağılımı yer almaktadır. Grafikte çocukların yaş dağılımı incelendiğinde, 5-11 yaş grubu 13.220 çocuk ile en kalabalık grup olarak öne çıkmaktadır. Yaş ilerledikçe çocuk sayısında azalma gözlenirken, 12-14 yaş grubu 6.092, 15-17 yaş grubu ise 5.878 çocuk ile en küçük nüfusa sahip grubu oluşturmaktadır.



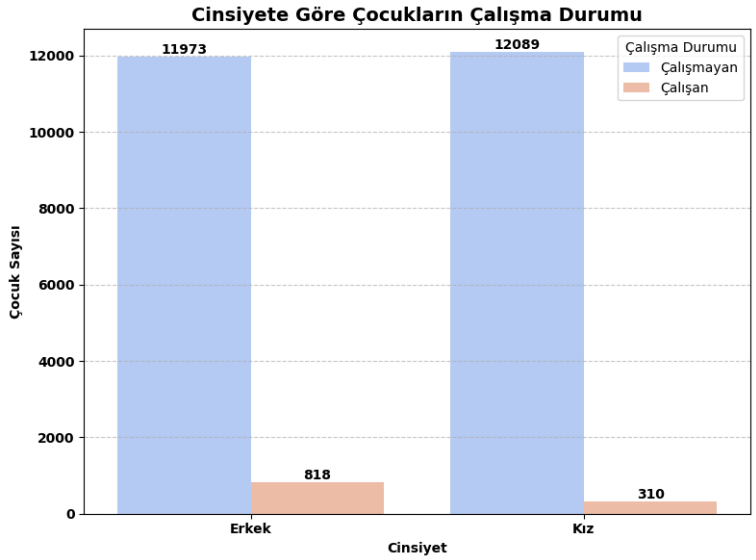
Şekil 2. Eğitime devam durumu.

Şekil 2'de görüldüğü üzere, 22.014 çocuk eğitimine devam ederken, 3.176 çocuk ise eğitimine devam etmemektedir. Eğitime devam etmeyen veya edemeyen çocuklar, ekonomik ya da ailevi nedenlerden ötürü risk altında olup, bunlara yönelik destek politikaları büyük önem taşımaktadır. Eğitime devam durumu dağılımı analiz edildiğinde, çocukların büyük bir kısmının (%87,4) eğitime devam ettiği, ancak %12,6'lık bir kesimin eğitimden kopmuş olduğu görülmektedir. Bu durum, genel olarak eğitim sisteminin kapsayıcı olduğunu göstermekle birlikte, 3.176 çocuğun eğitime devam etmemesi, dikkate alınması gereken önemli bir sorundur. Eğitime devam etmeyen çocukların, çocuk işçiliği riski altında olduğu değerlendirilmektedir. Soruna çözüm için; burs programları, çeşitli ekonomik destekler ve okula yeniden geri dönüş projeleri benzeri çözüm önerileri geliştirilebilir.



Şekil 3. Çocukların çalışma durum dağılımı.

Şekil 3'te, çocukların çalışma durumu dağılımı gösterilmektedir. Grafikte, çocukların çoğunlukla çalışmadığı ancak %4,5'lik bir kesimin çalışmak zorunda kaldığı görülmektedir. Ekonomik yetersizlikler, ailevi sorumluluklar veya sosyal faktörler bu duruma sebep olabilir. Çocuk işçiliğinin önlenmesi için çeşitli önlemlerin alınması önem arz etmektedir. Çeşitli eğitim teşvikleri, aile destek programları ve denetim mekanizmalarının güçlendirilmesi benzeri alınacak önlemler kritik önemdedir.



Şekil 4. Cinsiyete göre çocukların çalışma durumu.

Şekil 4'te, çalışan çocukların çoğunluğunu erkekler (818 kişi) oluştururken, kızlar ise 310 kişi ile daha düşük bir orana sahiptir. Bu durum, erkek çocukların ekonomik faaliyetlere daha fazla katıldığını, kız çocuklarının ise ev içi sorumluluklara yönelebileceğini göstermektedir. Mikro verilerden elde edilen analiz sonuçları, cinsiyete dayalı farklılıkların çocuk işçiliğinde belirleyici olduğunu ortaya koymaktadır.

Tablo 2, dört farklı makine öğrenmesi modelinin sınıflandırma performansını gösteren metrikleri ifade etmektedir. Random Forest ve SVM, F1-Skoru açısından en yüksek performansı sergilerken, XGBoost ve CatBoost modelleri birbirlerine yakın sonuçlar vermiştir. Metriklerin her biri model performanslarını farklı açılardan değerlendirmektedir.

Tablo 2. Makine öğrenmesi modelleri ve metrikleri.

Modeller	Doğruluk (Accuracy)	Kesinlik (Precision)	Duyarlılık (Recall)	F1-Skoru (F1-Score)
Random Forest	0,761	0,953	0,761	0,830
XGBoost	0,757	0,954	0,757	0,828
SVM	0,760	0,953	0,760	0,830
CatBoost	0,759	0,954	0,759	0,829

Doğruluk, tüm tahminlerin doğru olma oranını ifade etmektedir. Yani, modelin toplam doğru tahminlerinin, toplam tahminlere oranı gösterilmektedir. Tablo 2'deki doğruluk değerleri %75,7 ile %76,1 arasında değişmektedir. Random Forest modeli, %76,1 ile en yüksek doğruluğa sahiptir. Ancak doğruluk, dengesiz veri kümelerinde yanıltıcı olabileceğinden, tek başına model performansını ölçmede yeterli bir metrik değildir.

Kesinlik, pozitif olarak tahmin edilenlerin ne kadarının gerçekten pozitif olduğunu gösterir. Yani, yanlış pozitifleri azaltma yeteneğini ölçer. Kesinlik, özellikle yanlış pozitiflerin maliyeti yüksek olduğunda önemli bir metriktir. Tabloya göre, tüm modellerin kesinlik değeri oldukça yüksektir (yaklaşık %95,3-95,4). Bu da modellerin doğru pozitif tahmin yapma konusunda başarılı olduklarını göstermektedir.

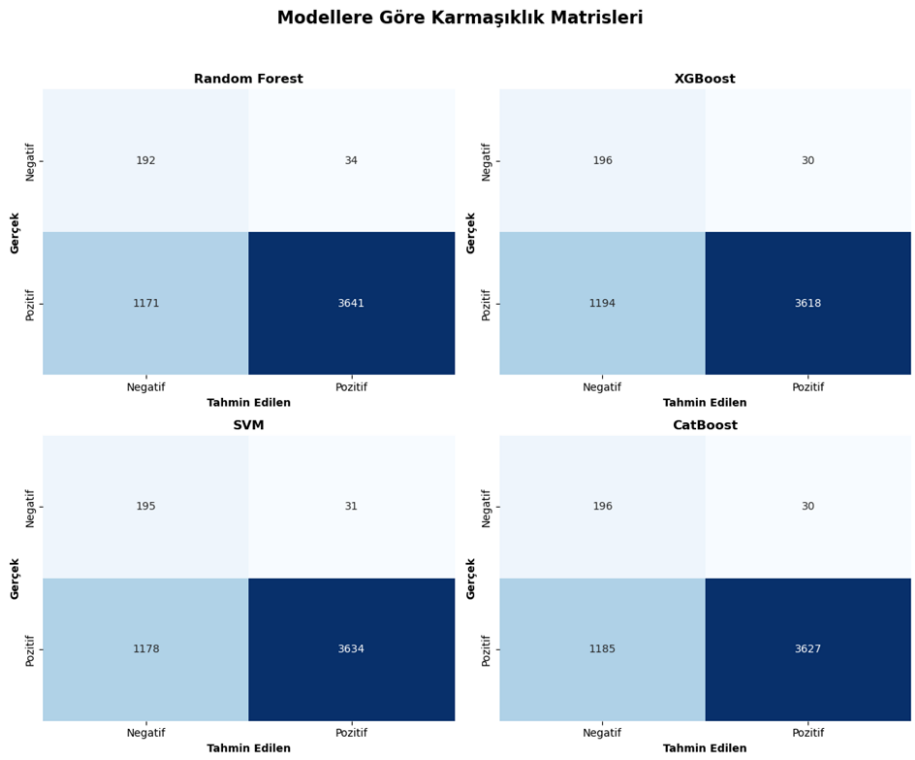
Duyarlılık, gerçek pozitiflerin ne kadarını doğru tahmin edebildiğimizi ifade eder. Yanlış negatifleri azaltmada önemli bir metriktir. Örneğin, bir hastalık teşhisi senaryosunda, duyarlılık düşükse birçok hasta yanlış şekilde sağlıklı olarak sınıflandırılabilir. Tabloda Random Forest ve SVM modelleri %76,1 ile en yüksek duyarlılığa sahiptir.

F1-Skoru, kesinlik ve duyarlılığın harmonik ortalamasıdır ve dengesiz veri kümelerinde genel başarıyı daha iyi temsil eder. Kesinlik ve duyarlılık arasında bir denge sağlamayı hedefler. Tabloya göre, Random Forest ve SVM modelleri %83 F1-Skoru ile en iyi performansı sergilemektedir. Bu, bu modellerin kesinlik ve duyarlılığı dengeli bir şekilde optimize ettiğini göstermektedir.

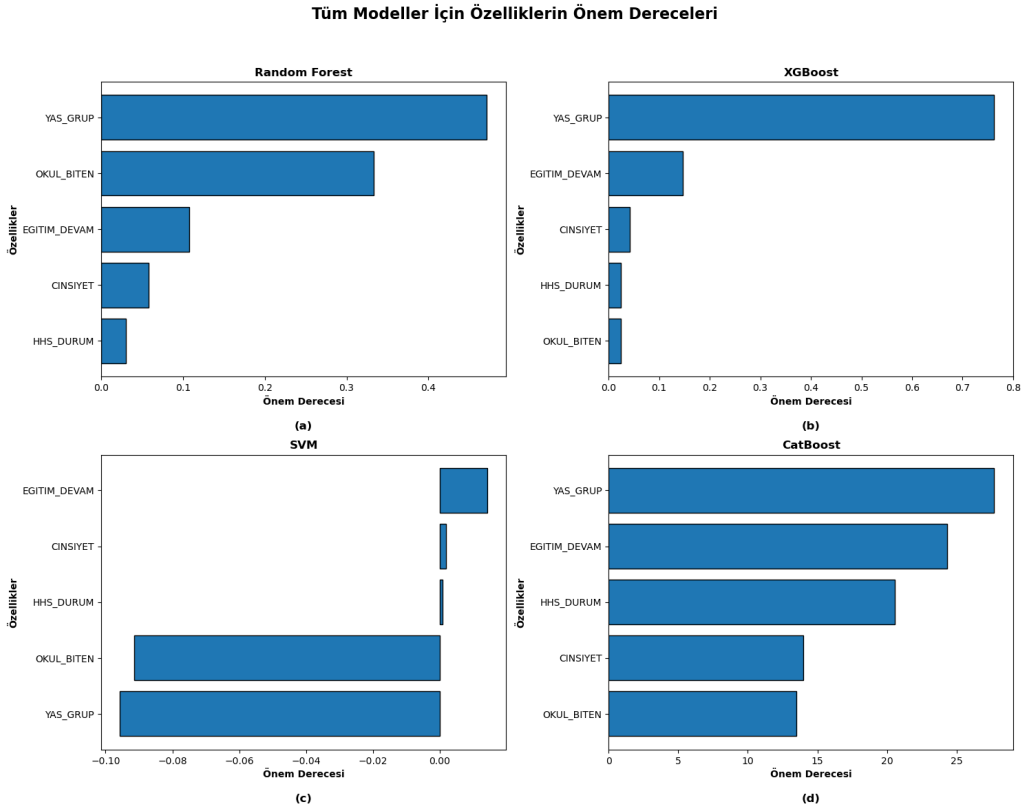
Şekil 5'te dört farklı makine öğrenmesi (a)Random Forest b)XGBoost c)SVM d)CastBoost) modelinin karmaşıklık matrislerini göstermektedir. Matrislerde, gerçek ve tahmin edilen sınıflar gösterilmektedir. Tüm modellerde Gerçek Pozitif sınıflarda yüksek başarı gözlenirken, yanlış negatifler (pozitif olan ancak negatif tahmin edilen) daha dikkat çekici bir farklılık oluşturmaktadır. Örneğin, Random Forest modeli 1171 yanlış negatif tahmin yaparken, XGBoost ve CatBoost modellerinde bu sayı 1194 ve 1185 olarak görülmektedir.

Öte yandan, Negatif Sınıfta yanlış pozitif tahminler oldukça düşük seviyelerdedir. XGBoost ve CatBoost modelleri, yanlış pozitif sayısını 30 ile minimize ederek diğer modellere kıyasla bu alanda daha iyi performans sergilemektedir. Genel olarak, tüm modeller dengeli sonuçlar üretmiş olsa da CatBoost ve XGBoost modelleri negatif sınıfta biraz daha iyi performans göstermektedir.

XGBoost, özellikle daha hızlı bir eğitim süreci sunması ve çok daha karmaşık veri yapılarında öne çıkması ile dikkat çekerken, CatBoost'un kategorik verileri otomatik olarak işlemesi, hiperparametre optimizasyonuna duyarlı olması ve GPU desteği sunması, özellikle kategorik değişkenlerin yoğun olduğu veri setlerinde daha başarılı olmasını sağlamaktadır. Bununla birlikte, CatBoost modeli, negatif sınıftaki tespitlerde daha az hata yaparak bu alanda öne çıkmıştır. XGBoost ise genelde daha düşük doğruluk oranlarına karşın, büyük veri setlerinde daha iyi genelleme yapma kapasitesine sahiptir.

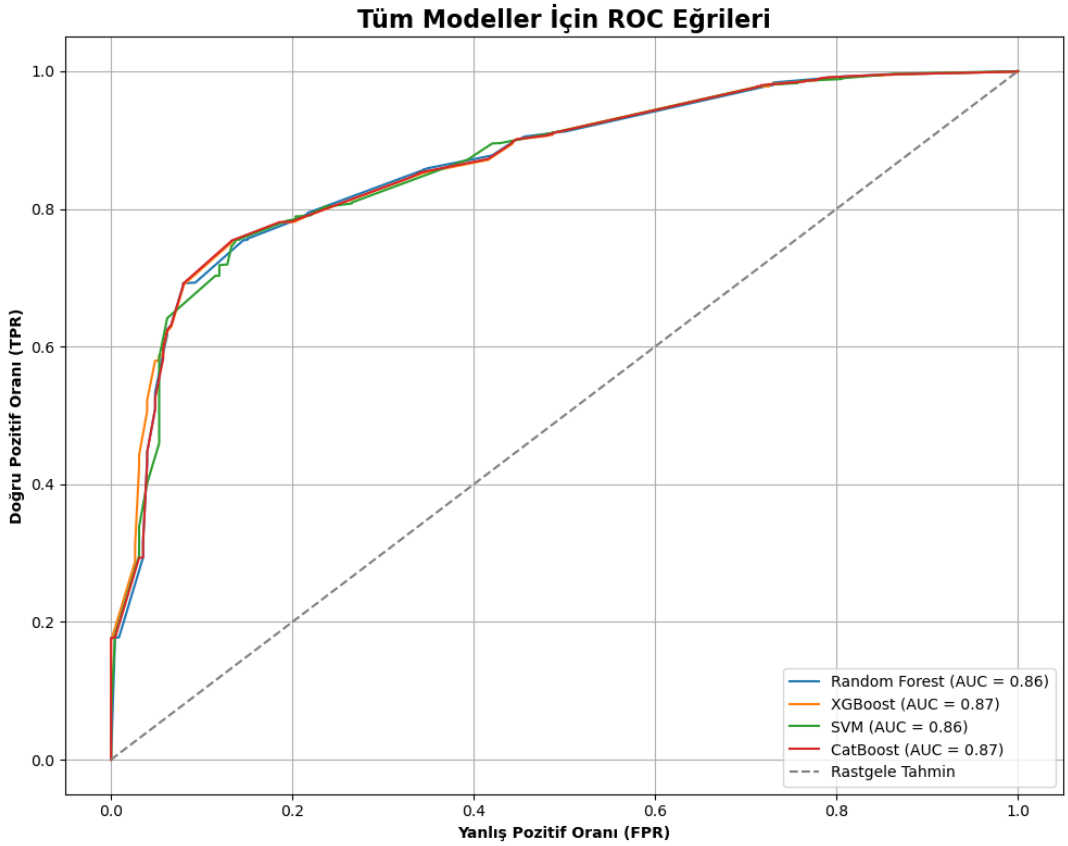


Şekil 5. Modellerin a)Random Forest b)XGBoost c)SVM d)CastBoost karmaşıklık matrisleri.



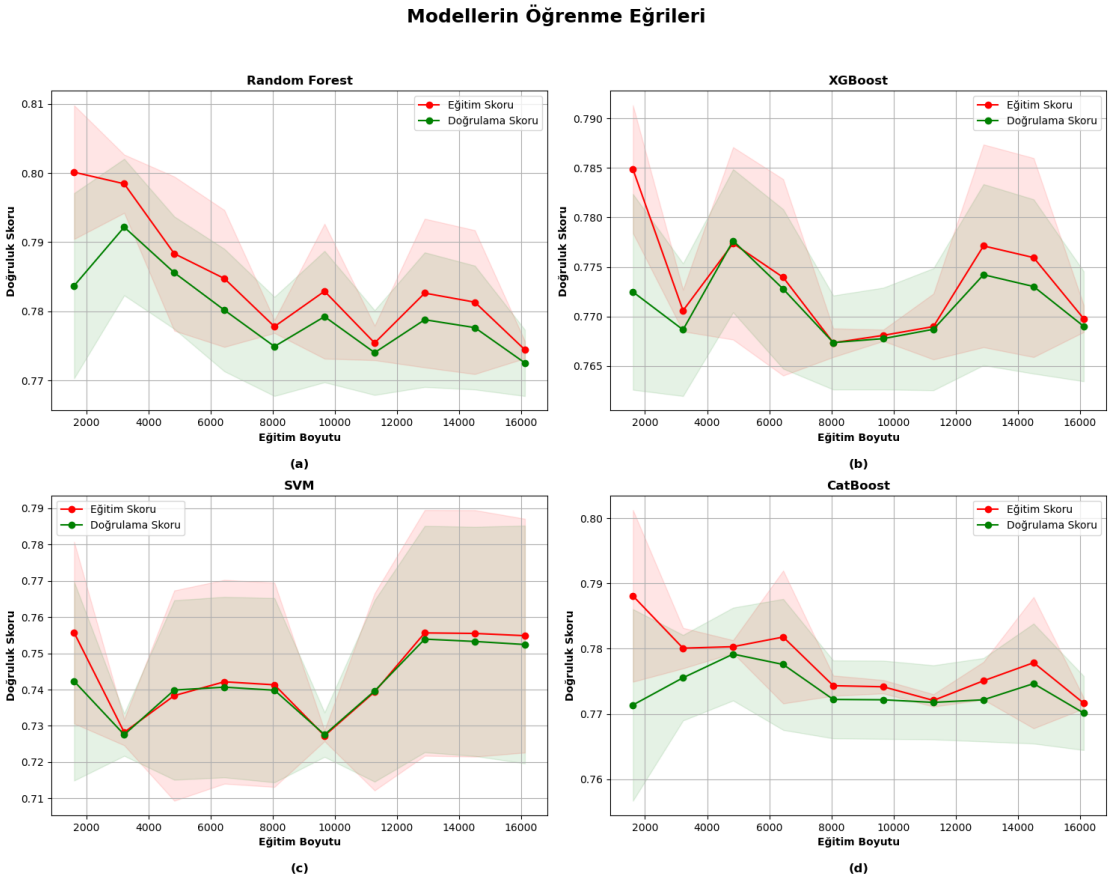
Şekil 6. Modellerin a)Random Forest b)XGBoost c)SVM d)CastBoost özellik önem dereceleri.

Şekil 6'da bütün modellerin özellik önem dereceleri gösterilmiştir. Şekil 6a'da yer alan Random Forest modelinde, YAS_GRUP değişkeni en kritik etkiye sahip değişkendir; onu EGITIM_DEVAM ve OKUL_BITEN değişkenleri takip etmektedir. Şekil 6b'deki XGBoost modelinde de YAS_GRUP açık ara en yüksek önem derecesinde olup; EGITIM_DEVAM ve OKUL_BITEN değişkenleri ise daha düşük oranlarda etkili olmuştur. Şekil 6c'deki SVM modelinde, EGITIM_DEVAM başlıca belirleyici etken olurken, OKUL_BITEN ve YAS_GRUP görece düşük fakat ters yönde bir etki ortaya koymuştur. Son olarak, Şekil 6d'de yer alan CatBoost modelinde YAS_GRUP değişkeni yine en yüksek önem dereceli değişken olarak ön plana çıkarken, EGITIM_DEVAM ise ikinci sırada yer almıştır. Elde edilen sonuçlar, yaş grubunun çocuk işgücü istihdamında kilit bir rol oynadığını ve eğitime devam etme durumunun da büyük ölçüde belirleyici bir faktör olduğunu göstermektedir. Modellerin farklı öğrenme stratejilerine bağlı olarak özelliklerin önem dereceleri de farklılık göstermektedir. Özellik önem analizlerinin dikkate alınması, tahmin performansını arttıran kritik bir unsurdur.



Şekil 7. Modellerin ROC eğrileri.

Şekil 7, bütün modeller için ROC eğrilerini ve AUC (Eğri Altındaki Alan) değerlerini göstermektedir. ROC eğrisi, modellerin sınıflandırma performansını Doğru Pozitif Oranı (TPR) ve Yanlış Pozitif Oranı (FPR) üzerinden değerlendirmektedir. XGBoost ve CatBoost modellerinin AUC değerlerinin 0.87 olduğu, Random Forest ve SVM modellerinin ise 0.86 değerinde olduğu görülmektedir. Tüm modellerin eğrileri birbirine oldukça yakın olup, sınıflandırma performansları tutarlı ve başarılıdır. Rastgele tahmin çizgisinin üzerinde konumlanan bu eğriler, modellerin rastgele sınıflandırmadan anlamlı ölçüde daha iyi performans sergilediğini ifade etmektedir.



Şekil 8. Modellerin a)Random Forest b)XGBoost c)SVM d)CastBoost öğrenme eğrileri.

Şekil 8, modellerin öğrenme eğrilerini eğitim boyutuna bağlı olarak göstermektedir. Öğrenme eğrileri; modellerin eğitim ve doğruluk skoru arasındaki ilişkiyi incelemekte ve modelin aşırı öğrenme (overfitting) veya yetersiz öğrenme (underfitting) durumlarını analiz etmeye olanak tanımaktadır. Şekil 8a'da Random Forest modeline ait öğrenme eğrisinde, eğitim skoru (kırmızı çizgi) ile doğrulama skorunun (yeşil çizgi) genel olarak birbiriyle yakın seyrettiği görülmektedir. Eğitim veri seti boyutu arttıkça skorlar arasında ciddi bir farklılık oluşmaması, modelin aşırı öğrenme eğiliminin düşük olduğunu belirtmektedir. Şekil 8b'de XGBoost modelinde de benzer biçimde eğitim ve doğrulama skorlarının birbirlerine yakın seyrettikleri görülmektedir. Ancak özellikle küçük veri boyutlarında, eğitim skoru ve doğrulama skoru arasında kısmen daha belirgin dalgalanmalar gözlemlenebilmektedir. Veri seti boyutu arttıkça bu dalgalanmalar azalmakta ve daha istikrarlı bir düzeye oturmaktadır. Şekil 8c'de SVM modelinin öğrenme eğrisinde, düşük veri boyutlarında eğitim skoru ile doğrulama skoru arasında zaman zaman daha keskin farklılıklar oluştuğu izlenmektedir. Bu durum, modelin başlangıç aşamasında veriye uyum sağlarken dalgalanma yaşadığını göstermektedir. Ancak veri miktarı arttıkça, eğitim ve doğrulama skorları daha dengeli bir hâle gelmekte ve skorlar arasındaki makas daralmaktadır. Şekil 8d'de CatBoost modelinde hem eğitim hem de doğrulama skorlarının görece stabil bir şekilde artış gösterdiği görülmektedir. Modellerin eğitim skorları ve doğrulama skorları arasında küçük farklılıklar görünmektedir. Bu durum da modellerin aşırı uyum (overfitting) sorunu yaşamadığını göstermektedir. Kesişen eğriler, modellerin eğitim verisi ve doğrulama verisi üzerinde benzer performanslar sergileyerek, genel anlamda doğru genellemeler yaptıklarına işaret etmektedir.

6. Sonuç

Makine öğrenimi modellerinin, karmaşık ilişkileri ve etkileşimleri anlamlandırma yetenekleri, politika yapıcılar ve akademisyenler için büyük fırsatlar barındırmaktadır. Yapılan çalışma, Türkiye'de çocuk işçiliğinin nedenlerini ve etkilerini analiz etmek üzerine, makine öğrenimi modellerinin öngörü potansiyelini değerlendiren

bir araştırma niteliği taşımaktadır. Analizlerde kullanılan veri seti, TÜİK 2019 Çocuk İşgücü Araştırması mikro verilerinden oluşmakta olup, 5-17 yaş arası çocukların istihdam, eğitim durumu, demografik özellikleri ve ailevi koşulları gibi çok boyutlu özelliklerini kapsamaktadır.

Resmi bir veri kaynağının kullanılması, elde edilen sonuçların doğruluğunu ve geçerliliğini artırırken, Random Forest, SVM, XGBoost ve Gradient Boosting modelleri ile yapılan analizler, veri odaklı stratejik politika geliştirme süreçlerine önemli katkılar sağlamaktadır. F1 Skoru gibi metriklerin kapsamlı bir şekilde değerlendirilmesi, modellerin performansını dengesiz veri kümelerinde bile etkili bir şekilde ölçebilmiştir. Çalışma bulguları hem akademik literatüre hem de politika yapıcıların veri temelli kararlar almasına yönelik kaynak sağlamaktadır.

Analiz sonucunda, çocukların yaş grubunun çalışma durumunu belirlemede en etkili değişken olduğu saptanmıştır. 5-11 yaş grubundaki çocuklar genellikle eğitimde kalırken, 15-17 yaş grubundaki çocukların istihdam oranları ise daha yüksektir. Cinsiyet bazında yapılan incelemede, erkek çocukların çalışma oranlarının kız çocuklara göre daha fazla olduğu tespit edilmiştir. Eğitime devam eden çocukların istihdama katılma olasılığının önemli ölçüte azaldığı saptanmıştır. Elde edilen bulgular, eğitim politikalarının çocuk işçiliği ile mücadelede kritik bir rol oynadığını ortaya koymuştur. Çocuk işgücü ile mücadelede bölgesel farklılıkları dikkate alan etkin ve yöreye özgü stratejiler geliştirilebilir ve eğitim destekli sosyal yardım projelerine odaklanılabilir. Çocuk işgücü kullanımının yaygın olduğu tarım, tekstil ve hizmet gibi sektörlerde etkin denetim mekanizmaları kurulması önemlidir.

Makine öğrenimi modellerinin performans karşılaştırmasında, Random Forest ve SVM modelleri, F1 skoru açısından en başarılı algoritmalar olarak ön plana çıkmaktadır. XGBoost ve CatBoost modelleri ise doğruluk oranlarında tutarlı performanslar sergilemiş, fakat daha düşük F1 skorları ile nispeten daha başarısız olmuştur. Yapılan analizlerde yaş grubu ve eğitim durumu değişkenlerinin tahmin modelleri üzerindeki etkisi vurgulanmıştır.

Sonuç olarak, bu çalışma, makine öğrenimi modellerinin sosyal bilimler alanındaki potansiyelini vurgulayarak, akademik literatüre ve politika yapıcılara veri odaklı stratejiler geliştirme konusunda değerli katkılar sağlamaktadır. Gelecekte yapılacak benzer çalışmalar, daha geniş veri setleri ve hibrit modeller kullanılarak yapılabilir. Yapılan çalışma hem mevcut sorunların tespitinde hem de gelecekteki iyileştirme çabalarına yönelik veri odaklı yaklaşımların geliştirilmesi açısından rehber niteliğindedir.

Kaynaklar

[1] T.C. Aile Çalışma ve Sosyal Hizmetler Bakanlığı. Çocuk İşçiliği ve Eğitim Öğretmen El Kitabı. Ankara: Genel Yayın No: 72, 2018. ISBN: 978-605-7888-00-6.

[2] Türkiye İstatistik Kurumu. Türkiye Çocuk İşgücü Mikro Veri Seti. Ankara, 2019.

[3] Eşidir KA, Gür YE. Yapay sinir ağları ile Türkiye plastik sektörü ithalat tahmini: 2023 yılı Nisan-Aralık ayları. Akad Hassasiyetler. 2023;10(23):91-114.

[4] Zhu X, Sawhney R, Upreti G. Determinates of employee voluntary turnover and forecasting in departments: A case study. Stud Eng Technol. 2016;3(1):64-73.

[5] Ji H. Robustness analysis on stock market prediction method. Highl Bus Econ Manag 2023;21:791-801.

[6] Bae CY, Im Y, Lee J, Park C, Kim M, Kwon HU, Kim J. Comparison of biological age prediction models using clinical biomarkers commonly measured in clinical practice settings: AI techniques vs. traditional statistical methods. Front Anal Sci. 2021;1.

[7] Pakarinen O, Karsikas M, Reito A, Lainiala O, Neuvonen P, Eskelinen A. Prediction model for an early revision for dislocation after primary total hip arthroplasty. PLOS ONE. 2022;17(9):e0274384.

[8] Speer AB. Empirical attrition modelling and discrimination: Balancing validity and group differences. Hum Resour Manag J. 2021;34(1):1-19.

[9] Wu Y. Job embeddedness review: Presentation, measurement and development. Adv Econ Manag Pol Sci 2023;47(1):169-174.

[10] Çöpoğlu M. Türkiye'de Çocuk İşçiliği. Iğd Üniv Sosyal Bilim Derg. 2018;(14):357-398.

[11] Gül T, Öztürk M. Çocuk İşçiliğinin Nedenleri Üzerine Kavramsal Bir Çalışma. Süleyman Demirel Üniv Sosyal Bilim Enst Derg. 2020;(37):130-148.

[12] Tosunoğlu E, Yılmaz R, Özeren E, Sağlam Z. Eğitimde Makine Öğrenmesi: Araştırmalardaki Güncel Eğilimler Üzerine İnceleme. Ahmet Keleşoğlu Eğ Fak Derg. 2021;3(2):178-199.

[13] Eriş Dereli B. Çocuk istihdamını etkileyen faktörler. İstanbul Tic Üniv Sosyal Bilim Derg. 2021;20(42):1505-1519.

[14] Kömüryakan F, Çağlayan E. Türkiye'de Çocuk İşçiliğinin Simetrik ve Asimetrik Modeller İle Analizi: Logit, Probit, Log-Log ve Clog-Log. Hacettepe Üniv İktis İdari Bilim Fak Derg. 2022;40(4):776-798.

[15] Yardımcıoğlu D. Çocuk İşçiliğini Önlemeye Yönelik Yeni Bir Adım: Avrupa Parlamentosu ve Konseyi'nin Kurumsal Sürdürülebilirlik Durum Tespitine Dair Direktif Önerisi. Dicle Üniv Hukuk Fak Derg. 2023;28(48):225-276.

[16] Özdemir NE, Akyiğit Albayrak E, Korkutan M. Türkiye'de çocuk işçiliği: sağlık ve sosyal yaşam koşulları bağlamında bir değerlendirme. Sosyal Bilim Akademi Derg. 2024;7(1):44-66.
[17] Suthaharan S. Big data classification: Problems and challenges in network intrusion prediction with machine learning. ACM SIGMETRICS Perform Eval Rev. 2014;41(4):70-73.
[18] El Naqa I, Murphy MJ. What is machine learning? In: Machine Learning in Radiation Oncology. Springer; 2015:3-11.
[19] Eşidir KA, Gür YE. Multilayer Perceptron (MLP) ile Türkiye İşlenmemiş Alüminyum Sektörü İthalat Tahmini: 2023 Yılı Nisan-Aralık Ayları Dönemi Üzerine Bir İnceleme. Erciyes Üniv İktis ve İdari Bilim Fak Derg. 2024;(68):57-64.
[20] Gür YE. Development and application of machine learning models in US consumer price index forecasting: Analysis of a hybrid approach. Data Sci Finance Econ. 2024;4(4):469-513.
[21] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Computer Vision – ECCV 2014. Springer; 2014:818-833.
[22] Gür YE. Stock price forecasting using machine learning and deep learning algorithms: A case study for the aviation industry. Fırat Üniv Müh Bilim Derg. 2024; 36(1): 25-34.
[23] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016.
[24] Gür YE. Forecasting the euro exchange rate using deep learning algorithms and machine learning algorithms. İstanbul Tic Üniv Sosyal Bilim Derg. 2024; 23(49): 1435-1456.
[25] Wang L, Wang X, Chen A, Jin X, Che H. Prediction of Type 2 Diabetes Risk and its effect evaluation based on the XGBoost model. Healthcare. 2020; 8(3): 247.
[26] Liang W, Luo S, Zhao G, Wu H. Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms. Mathematics. 2020; 8(5): 765.
[27] Oguine OC, Oguine MB. Comparative analysis and forecasting on the death rate of COVID-19 patients in Nigeria using random forest and multinomial Bayesian epidemiological models. J Clin Case Stud Rev Rep. 2021; 1-7.
[28] Thi Mai H. Van, Hoang Trinh S, Ly H. B. Enhancing compressive strength prediction of roller compacted concrete using machine learning techniques. Meas J Int Meas Confed. 2023; 218(February): 113196.
[29] Ayhan S, Erdoğmuş Ş. Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. Eskişehir Osmangazi Üniv İktis ve İdari Bilim Derg. 2014; 9(1): 175-201.
[30] Cao T, She D, Zhang X, Yang Z. Understanding the influencing factors and mechanisms (land use changes and check dams) controlling changes in the soil organic carbon of typical loess watersheds in China. Land Degrad Dev. 2022; 33(16): 3150-3162.
[31] Wen J, Zhang Y, Yang G, He Z, Zhang W. Path loss prediction based on machine learning methods for aircraft cabin environments. IEEE Access. 2019; 7: 159251-159261.
[32] Zhang Z, Xin Q, Li W. Machine learning-based modeling of vegetation leaf area index and gross primary productivity across North America and comparison with a process-based model. J Adv Model Earth Syst. 2021; 13(10).
[33] Eşidir KA. Türkiye'nin kimyasal madde ithalatının gelecek tahmini: Makine öğrenmesi ve topluluk öğrenme yöntemleri performans analizi. Fırat Üniv Sosyal Bilim Derg. 2025; 35(1): 261-278.